FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Optimal Stopping Problems with Expectation Constraints

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der
Fakultät für Mathematik und Informatik
der Friedrich-Schiller-Universität Jena

von M.Sc. Maike Klein
geboren am 12.09.1989 in Bad Honnef

Gutachter

1. Prof. Dr. Stefan Ankirchner, Friedrich-Schiller-Universität Jena

2. Prof. Dr. Mikhail Urusov, Universität Duisburg-Essen

3. Prof. Dr. Goran Peskir, The University of Manchester, UK

Tag der öffentlichen Verteidigung: 24.09.2018

"Solving a problem for which you know there's an answer is like climbing a mountain with a guide, along a trail someone else has laid. In mathematics, the truth is somewhere out there in a place no one knows, beyond all the beaten paths. And it's not always at the top of the mountain. It might be in a crack on the smoothest cliff or somewhere deep in the valley."

Yoko Ogawa

"Life isn't about waiting for the storm to pass,
it's about learning how to dance in the rain."

Vivian Greene

Acknowledgments

First of all, I would like to express my sincere gratitude to my advisor Stefan Ankirchner for the guidance in all phases of my work, for many fruitful discussions, his constant support and motivation. His enthusiasm and passion for mathematics have always encouraged me to face difficult mathematical problems. I have benefited a lot from working with him.

I am grateful to Goran Peskir for hosting me for two months and for getting me acquainted with other researchers. I had a wonderful and enriching time in Manchester.

I would like to thank Mikhail Urusov for helpful comments on my research and for agreeing to review this thesis.

Moreover, many thanks to Thomas Kruse and Nabil Kazi-Tani for stimulating discussions and interesting questions. Many results of this thesis arose from our joint work with Stefan Ankirchner.

For an atmosphere of friendship and appreciation, for a lively and joyful working environment, for mathematical discussions as well as for unforgettable joint times, I am grateful to Alexandra, Daniel, Lena, Linh, Markus, Max, Nadine, Nicole, Robert and Stefan. Many thanks to Robert for valuable suggestions that contributed to improve my results. Dear Alexandra, Linh and Markus, sharing the office with you was a great pleasure. Thank you for many moments of laughter and insight.

I owe my deepest thanks to my family and friends, especially to my parents for their love, advice and support; to my mum for her care and strength; to my dad for his reliability and for sharing his love of music with me. To all, who have supported me during my personal storm, thank you for showing me how to dance in the rain.

Financial support by the DAAD for my research stay at the University of Manchester is gratefully acknowledged.

Zusammenfassung

In der vorliegenden Arbeit werden optimale Stoppprobleme mit Restriktionen an ein durchschnittliches Kostenfunktional der Stoppzeit untersucht. Das Ziel ist es, sowohl die Menge der Stoppzeiten zu reduzieren als auch eine partielle Differentialgleichung für die Wertfunktion herzuleiten.

Ist der zu stoppende Prozess ein zeithomogener Itô-Prozess, können wir das Stoppproblem durch Erweiterung des Zustandsraumes in ein optimales Kontrollproblem ohne Nebenbedingungen überführen und so ein dynamisches Programmierungsprinzip erhalten. Die Wertfunktion wird durch eine elliptische nichtlineare partielle Differentialgleichung zweiter Ordnung charakterisiert. Wir beweisen ein klassisches Verifikationstheorem und wenden es auf mehrere Beispiele an.

Des Weiteren betrachten wir optimale Stoppprobleme für eindimensionale reguläre stetige starke Markovprozesse, wobei der Erwartungswert der Stoppzeiten beschränkt ist. Wir zeigen, dass es ausreichend ist, Stoppzeiten zu betrachten, sodass die Verteilung des Prozesses zur Stoppzeit einer gewichteten Summe von 3 Diracmaßen entspricht. Der Beweis basiert auf Ergebnissen zur Skorokhod Einbettung und überführt das Stoppproblem in ein lineares Optimierungsproblem über einer konvexen Menge von Wahrscheinlichkeitsmaßen.

Die Ergebnisse werden in der Analyse eines sequentiellen Testproblems angewendet. Wir zeigen, dass die optimalen Stoppzeiten in diesem Problem durch höchstens zwei aufeinanderfolgende Austrittszeiten gegeben sind.

Abschließend untersuchen wir mit Hilfe der Theorie der Tchebycheffsysteme, unter welchen Voraussetzungen die Menge der Stoppzeiten auf erste Austrittszeiten aus Intervallen reduziert werden kann. Die Verteilung des Prozesses zur Stoppzeit ist in diesem Fall eine gewichtete Summe von 2 Diracmaßen.

Abstract

In this thesis we investigate optimal stopping problems with expectation cost constraints. We focus on reducing the set of stopping times as well as on deriving a partial differential equation for the value function.

If the process to stop is a time-homogeneous Itô-process, we show, by introducing a new state variable, that one can transform the problem into an unconstrained control problem and hence obtain a dynamic programming principle. We characterize the value function in terms of the dynamic programming equation, which turns out to be an elliptic, fully non-linear partial differential equation of second order. In addition, we prove a classical verification theorem and apply it to several examples.

Furthermore, we consider the problem of optimally stopping a one-dimensional regular continuous strong Markov process with a stopping time satisfying an expectation constraint. We show that it is sufficient to consider only stopping times such that the law of the process at the stopping time is a weighted sum of 3 Dirac measures. The proof uses results on Skorokhod embeddings in order to reduce the stopping problem to a linear optimization problem over a convex set of probability measures.

We apply the results to analyze a sequential testing problem and show that in this problem the optimal stopping times are given by at most two consecutive exit times of intervals.

Finally, using the theory of Tchebycheff systems we examine when we can reduce the set of stopping times in the constrained problem to first exit times of intervals. In this case, the law of the process at the stopping time is a weighted sum of at most 2 Dirac measures.

Contents

I. Introduction

In an optimal stopping problem one chooses an optimal point in time to perform a particular action in order to maximize an expected payoff or to minimize expected costs. Since future rewards are uncertain and stopping decisions are non-reversible, the decision maker has to weigh possible future gains and losses against receiving a secure reward today.

The study of optimal stopping problems was stimulated by a work of Wald [69] in 1947. He considered a sequential testing problem, where one has to decide between two hypotheses based on independent observations. Wald determined the optimal time to stop collecting data when taking in expectation as few samples as possible while simultaneously reducing the probability of an incorrect decision. He showed that his method results in a smaller average number of samples than methods with fixed sample size and the same error probabilities.

Many optimization problems in economics, financial mathematics or statistics can be interpreted as optimal stopping problems. Consider for example the so-called secretary problem. Here the decision maker chooses a stopping rule that maximizes the probability of selecting the best offer among a sequence of offers (see [19, Chapter 3]). In financial mathematics determining the optimal exercise date for an American option as well as pricing other financial contracts are optimal stopping problems (see e.g. [35] and [28]). Production scheduling in investment models with market entry and exit decisions can also be modeled as optimal stopping problems, cf. [23].

For an introduction to optimal stopping problems, applications, an historical overview and references we refer to the books [55] and [63].

Now consider an optimal stopping problem with finite or infinite time horizon. If the process to stop is *Markovian*, at every time $t \in [0, \infty)$ the future behavior of the process X only depends on the present state X_t and not on the whole path until time t. Thus, for deciding whether to stop at time t we only have to take into account the value of X_t. Hence, the state space of the underlying process can be split into two regions: In the stopping region it is optimal to stop, whereas in the continuation region we do not stop. An optimal stopping time is the first time the process enters the stopping region. Therefore, if X is a one-dimensional continuous Markov process, the law of X at the optimal stopping time is a weighted sum of at most 2 Dirac measures. The stopping region can be identified as the free boundary of a partial differential equation (PDE), see e.g. [44] and Section 8 in [55]. For an infinite or finite time horizon optimal stopping problem the PDE in variational form for the value function V is given by

$$\min\big\{ -\mathcal{L}V(x), V(x) - f(x)\big\} = 0, \qquad \text{and} \qquad \min\big\{ -(\partial_t + \mathcal{L})V(t,x), V(t,x) - f(x)\big\} = 0,$$

respectively, where $\mathcal{L}$ denotes the generator of the underlying Markov process and f is the payoff function (see e.g. [55, Section 8] or [67, Chapter 4]).

The first part of this thesis (Chapter II) analyzes optimal stopping problems with expectation cost constraints. This problem captures situations with an *average* time/cost limit for any stopping rule. As a special case we obtain the stopping problem over all stopping times τ with expectation constraint $\mathbb{E}[\tau] \leq T$. Whenever a stopping rule τ is applied *repeatedly* and *independently* of the previous stopping times, then an average constraint on the costs H_τ seems to be more appropriate than a sharp constraint of the form $H_\tau \leq T$, a.s. For example, suppose you are working in a human resources department of a big company and you have to decide when to stop searching for a new employee, then it is more likely that you impose an average constraint on your searching time than just a sharp upper bound.

We aim at characterizing the value function of the optimal stopping problem in terms of a dynamic programming equation and at proving a classical verification theorem.

Let $(X_t)_{t\in[0,\infty)}$ be an n-dimensional stochastic state process that satisfies a time-homogeneous stochastic differential equation driven by a d-dimensional Brownian motion W. Denote by $(\mathcal{F}_t)$ the filtration that is generated by W and extended by null sets so as to satisfy the usual conditions. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a payoff function and $H_t = \int_0^t h(X_s)ds$ be a strictly increasing cost process, where $h\colon \mathbb{R}^n \to (0,\infty)$. We denote by $\mathcal{S}(T)$ the set of $(\mathcal{F}_t)$-stopping times τ satisfying the constraint $\mathbb{E}[H_\tau] \le T \in [0,\infty)$. We consider the optimal stopping problem

$$\text{maximize } \mathbb{E}\big[f(X_\tau)\big] \quad \text{subject to } \tau \in \mathcal{S}(T). \tag{I.1}$$

By choosing $h(x) = 1$ for all $x \in \mathbb{R}^n$, we obtain the stopping problem with expectation constraint $\mathbb{E}[H_\tau] = \mathbb{E}[\tau] \le T$.

What makes the stopping problem (I.1) difficult is that there is no simple dependence of the constraint on time. The expectation constraint has to be turned into a path-dependent constraint. A first attempt to eliminate the constraint is to follow a Lagrange approach and to consider, for every $\lambda > 0$, the unconstrained stopping problem

$$w(\lambda) = \sup\big\{\mathbb{E}[f(X_\tau) - \lambda H_\tau]\colon \tau \text{ stopping time with } \mathbb{E}[H_\tau] < \infty\big\}. \tag{I.2}$$

Notice that (I.2) is an infinite time horizon stopping problem that does not involve a discount factor. Therefore, it is often impossible to characterize w as the unique solution of a dynamic programming equation (cf. Section II.6). Disregard this for a moment and assume that we can identify for some $\bar\lambda > 0$ an optimal stopping time $\tau^*\big(\bar\lambda\big)$ for $w(\bar\lambda)$ with $\mathbb{E}\big[H_{\tau^*(\bar\lambda)}\big] = T$, then the stopping time $\tau^*(\bar\lambda)$ is optimal for the original problem (I.1), because

$$\begin{aligned}\mathbb{E}\big[f\big(X_{\tau^*(\bar\lambda)}\big)\big] \le \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}\big[f(X_\tau)\big] \le \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}\big[f(X_\tau) - \bar\lambda H_\tau\big] + \bar\lambda T \le w\big(\bar\lambda\big) + \bar\lambda T \\ = \mathbb{E}\big[f\big(X_{\tau^*(\bar\lambda)}\big) - \bar\lambda H_{\tau^*(\bar\lambda)}\big] + \bar\lambda T = \mathbb{E}\big[f\big(X_{\tau^*(\bar\lambda)}\big)\big].\end{aligned}$$

Moreover, if for every $\lambda > 0$ there exists an optimal stopping time $\tau^*(\lambda)$ for (I.2), if w is absolutely continuous with $\frac{\partial w}{\partial\lambda}(\lambda) = -\mathbb{E}[H_{\tau^*(\lambda)}]$ and if there exists $\bar\lambda$ such that $\frac{\partial w}{\partial\lambda}(\bar\lambda) = -T$, then the stopping time $\tau^*(\bar\lambda)$ satisfies $\mathbb{E}\big[H_{\tau^*(\bar\lambda)}\big] = T$. It can happen, however, that the function w is not absolutely continuous (see Section II.6 for an example). Even if w is differentiable, then it can be involved and error-prone to invert the derivative $\frac{\partial w}{\partial\lambda}$ and to determine the appropriate Lagrange multiplier $\bar\lambda$.

In Chapter II, which is based on [4], we propose a new approach for solving stopping problems of the type (I.1). Our basic idea is to extend the state space by the conditional expectation process $M_t = \mathbb{E}[H_\tau \,|\, \mathcal{F}_t]$, $t \in [0,\infty)$. Consequently, the expectation constraint for the costs H_τ is turned into an initial value for the new state M. Moreover, at every time t the expected remaining costs are given by $M_t - H_t$. Thus, if the conditional expectation process falls below H_t, the remaining costs are 0 and hence, the stopping time can be identified as the first time $t \in [0,\infty)$ such that $M_t \le H_t$. Assuming a Brownian set-up, the predictable representation property allows to interpret the new state variable as a martingale with controlled diffusion coefficient. Moreover, M solves

$$M_t = m + \int_0^t \mathbb{1}_{\{M_s > H_s\}} \alpha_s \cdot dW_s$$

for a suitable $\alpha \in L^2_{loc}(W)$ and $m = \mathbb{E}[H_\tau]$. Using a one-to-one correspondence between stopping times satisfying an expectation cost constraint and suitable controlled martingales, one can thus transform the stopping problem (I.1) into an unconstrained problem with a controlled state and time horizon. The advantage of the transformed problem is that it allows to formulate a dynamic programming principle (DPP). The DPP can be interpreted as follows: An optimal control for an optimization problem, where the controlled process $(Z_t)_{t\in[0,\infty)}$ starts in z at time 0, is also optimal in the problem, where the process starts in Z_θ at a stopping time θ (see [9, p.83]).

With a DPP at hand, we can characterize the value function

$$V(T,x) = \sup\big\{\mathbb{E}[f(X_\tau)] \colon \tau \in \mathcal{S}(T) \text{ and } X_0 = x\big\}$$

as a solution of the dynamic programming equation (DPE), which is the infinitesimal counterpart of the DPP. Conversely, in a classical verification theorem we establish conditions which guarantee that a classical solution to the DPE is indeed the value function of the optimal stopping problem. In order to obtain a verification theorem, we consider the auxiliary stopping problem

$$U(T,x) = \sup\big\{\mathbb{E}[f(X_\tau)] \colon \tau \text{ stopping time with } \mathbb{E}[H_\tau] = T \text{ and } X_0 = x\big\}. \tag{I.3}$$

Observe that if $U(T,x)$ is increasing in T on some interval $[0,S)$, $S\in(0,\infty)$, then the value function of the auxiliary and original problem coincide on $[0,S)$. Moreover, if U is concave in T, then one can show that $V(T,x) = U\big(T\wedge\widetilde{T}(x),x\big)$, where $\widetilde{T}(x) = \inf\big\{t\in[0,\infty)\colon U(t,x) < \sup_{s\le t} U(s,x)\big\}$. The concavity of U will be a consequence of the verification theorem for the problem (I.3). The DPE for U turns out to be the partial differential equation

$$h(x)U_T(T,x) - \mathcal{L}U(T,x) + \frac{\big|\sigma^\top(x)\cdot\nabla_x U_T(T,x)\big|^2}{2U_{TT}(T,x)} = 0, \tag{I.4}$$

with initial condition $U(0,x) = f(x)$; here $\mathcal{L}$ is the generator of X and σ its diffusion matrix. We give sufficient conditions for the value function U to be a solution of (I.4). Moreover, we provide a verification theorem that allows to verify whether a solution u of (I.4) coincides with U and if u equals U, then we obtain an optimal control for the value function U. Using the one-to-one correspondence between stopping times satisfying an expectation cost constraint and suitable controlled martingales, we derive an optimal stopping time for (I.3). Since V is fully determined by U, this further allows to identify an optimal stopping time for the original problem (I.1).

The idea to extend the state space by a conditional expectation process in order to make a constraint more tangible can be already found in the control literature. Bouchard, Elie and Touzi [13] consider the problem of attaining a possibly stochastic target with a given probability. They extend the state space by a conditional probability process in order to reduce the problem to a standard stochastic target problem. Bouchard, Elie and Imbert [12] deal with an optimal control problem with constraints on the moment or a certain probability of the terminal value of the controlled process. They obtain an optimal control problem with a stochastic target constraint by extending the state space with a conditional expectation process. Bouchard and Nutz [14] examine an optimal control problem with generalized state constraints and prove a dynamic programming principle. By a suitable extension of the state space, the state constraint can be turned into an expectation constraint for the terminal value of the controlled process. Augmenting the state space of the new problem by auxiliary martingales, the marginal expectation constraint is formulated dynamically and allows to prove a weak dynamic programming principle. The value function of the optimal control problem with expectation constraint is characterized as a viscosity solution to the associated Hamilton-Jacobi-Bellman equation. Bokanowski, Picarelli and Zidani [11] reformulate a stochastic control problem with a global state constraint as a target problem by introducing a conditional expectation process as a new controlled variable. Article [73] derives a dynamic programming principle for optimal control problems on the canonical space with expectation constraints in a general non-Markovian framework. At a given time $t\in[0,T]$, $T\in(0,\infty)$, the authors maximize the expected reward over a set of controls ν that satisfy for all $s\in[t,T]$ the constraint $\mathbb{E}[g(s,X^{t,\nu})]\le m$, where $m\in\mathbb{R}$ and $X^{t,\nu}$ is the controlled process whose paths are fixed from time 0 to t. For the dynamic programming principle they introduce auxiliary supermartingales M with $M_s \ge g(s,X^{t,\nu})$, $s\in[t,T]$. These supermartingales comprise processes of the form $m + \int_0^\cdot \alpha_s dW_s$ for suitable $\alpha\in L^2(W)$, where W is a Brownian motion.

Independently of article [4], Miller [43] has obtained the PDE (I.4). The author uses the auxiliary stopping problem (I.3) for solving a time-inconsistent, but unconstrained stopping problem. For a one-dimensional Brownian motion the value function $U(T,x)$ is characterized as the unique viscosity solution of (I.4). In contrast, we focus on deriving a classical verification theorem. Bayraktar and Yao [7] provide a proof of the dynamic programming principle which we formulate in Section II.3.3 and characterize the value function of the stopping problem as the unique viscosity solution of the associated Hamilton-Jacobi-Bellman equation.

There are only few further articles in the literature that deal with stopping problems of the type (I.1). Kennedy [39] considers the problem of stopping a discrete time process with the constraint that the expectation of any stopping time is bounded by some given constant. He uses Lagrangian techniques for determining optimal stopping rules. Horiguchi [34] considers optimal stopping of a finite state process that, in addition, can be controlled with finitely many actions. Optimal stopping rules satisfying an expectation constraint are determined with mathematical programming techniques. Palczewski and Stettner [50] examine an undiscounted optimal stopping problem with infinite time horizon of the type (I.1) under the additional assumption that X is an ergodic, time-homogeneous weak Feller process. They state sufficient conditions guaranteeing that the set of stopping times can be restricted to those with bounded expectation. This boundary is in general not global but depends on the initial value of X.

Using a Lagrange approach, Makasu [42] obtains an upper bound for the value of an optimal stopping problem with expectation constraint. The gain is determined by a geometric Brownian motion and a coupled diffusion process whereas the transaction costs only depend on the diffusion process. Peskir [54] deals with a quickest detection problem and allows for expectation constraints. In Section 4 of [54] the expected positive or negative deviation of the stopping time from the time to detect is assumed to be bounded. The constrained problem can be fully solved by using the Lagrange approach. Similarly, the Lagrangian method is successfully employed in [52]. The problem of maximizing the mean of a stopped process subject to a variance constraint is first reduced to a mean-variance optimal stopping problem. The latter can be turned into a family of linear stopping problems by applying the Lagrangian approach once more.

In Chapter III, which is a revised version of [3], we investigate optimal stopping problems for one-dimensional state processes over stopping times τ satisfying the expectation constraint $\mathbb{E}[\tau] \leq T$. In contrast to Chapter II we also allow for processes that do not solve a time-homogeneous stochastic differential equations driven by a Brownian motion. More precisely, let $(Y_t)_{t\in[0,\infty)}$ be a one-dimensional regular continuous strong Markov process with respect to a right-continuous filtration $(\mathcal{F}_t)$. Let $f\colon \mathbb{R} \to \mathbb{R}$ be measurable. We consider the optimal stopping problem

$$\text{maximize } \mathbb{E}\big[f(Y_\tau)\big] \quad \text{subject to } \tau \in \mathcal{S}(T), \tag{I.5}$$

where $\mathcal{S}(T)$ is the set of $(\mathcal{F}_t)$-stopping times such that $\mathbb{E}[\tau] \leq T \in [0,\infty)$.

We show that for the stopping problem (I.5) it is sufficient to consider only stopping times τ such that the law of Y_τ is a weighted sum of at most 3 Dirac measures. Any such stopping time can be interpreted as a composition of exit times from intervals.

We also show that in general a reduction to weighted sums of 2 Dirac measures is not possible. In particular, one cannot split the state space into a deterministic stopping and continuation region. This is in contrast to stopping problems with a sharp bound on the stopping time and to stopping problems with infinite time horizon and discounting.

Our idea for proving a reduction to 3 Dirac measures is to rewrite the stopping problem (I.5) as a linear optimization problem over a set of probability measures.

For this purpose we use results on the Skorokhod embedding problem. Skorokhod [64] formulated and solved the so-called Skorokhod embedding problem in 1961, the English translation [65] appeared in 1965. Let μ be a centered probability measure on $\mathbb{R}$ with finite second moment and let W be a one-dimensional Brownian motion. We want to find an integrable stopping time τ such that W_τ is distributed according to μ. In this case, we say that τ embeds μ into W and write $W_\tau \sim \mu$. Since the last 50 years the original problem has been generalized and many solution methods have been developed. In particular, in 1971 Rost [60] gave necessary and sufficient conditions for the existence of embedding stopping times for general continuous Markov processes.

The article [47] by Obłój provides an overview of the developments until 2004. For applications in mathematical finance we refer to [31].

In Chapter III we focus on the solution method by Chacon and Walsh [17]. They use a one-to-one correspondence between integrable probability measures μ on $\mathbb{R}$ and potential functions u_μ, where

$$u_\mu(x) = -\int_{\mathbb{R}} |x-y|\,\mu(dy).$$

For a centered probability measure μ with finite second moment, their main idea is to construct an approximating sequence $(u_n)_{n\in\mathbb{N}}$ of potential functions such that $u_n(x) \downarrow u_\mu(x)$ for all $x \in \mathbb{R}$ as $n \to \infty$ and such that the corresponding probability measures μ_n, $n \in \mathbb{N}$, are discrete and have finitely many atoms. In particular, they construct a non-decreasing sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ such that $W_{\tau_n} \sim \mu_n$ and the stopping times τ_n are consecutive exit times of intervals. Moreover, it holds that $W_\tau \sim \mu$, where $\tau := \lim_{n\to\infty} \tau_n$, as well as $\mathbb{E}[\tau] = \int_{\mathbb{R}} x^2\mu(dx) < \infty$. We show in Section III.3 that the arguments of Chacon and Walsh also apply to a one-dimensional regular continuous strong Markov process Y and an integrable centered probability measure μ if the condition $\int_{\mathbb{R}} x^2\mu(dx) < \infty$ is replaced by an integrability condition which depends on the speed measure of Y. The expectation of an embedding stopping time τ for μ can be computed in terms of the measure μ and the speed measure of Y.

To reduce the optimal stopping problem (I.5) to a measure optimization problem, we characterize the set $\mathcal{A}(T)$ of probability measures that can be embedded into Y (see [2] and [32]) with stopping times having expectation smaller than or equal to T. The set $\mathcal{A}(T)$ consists of integrable probability measures satisfying moment constraints. We then apply the balayage method of Chacon and Walsh to construct for every $\mu \in \mathcal{A}(T)$ an embedding stopping time that is a limit of consecutive exit times. The balayage method brings along, for every embeddable distribution, an approximating sequence of probability measures that are weighted sums of finitely many Dirac measures. This allows to simplify the optimization problem by reducing the set of probability measures. Using a trick by Hoeffding [33] we can further reduce the set to those probability measures μ that are weighted sums of at most 3 Dirac measures. In addition, we use the Balayage method to construct an embedding stopping time for μ that is a composition of exit times.

In a second approach we exploit the linearity of the measure optimization problem. As for linear optimization problems in $\mathbb{R}^n$ with linear constraints, the maximal value of $\int_{\mathbb{R}} f(x)\mu(dx)$ over all measures $\mu \in \mathcal{A}(T)$ is attained by extreme points of $\mathcal{A}(T)$. Furthermore, the extreme points are contained in the set of weighted sums of Dirac measures with at most 3 mass points satisfying the moment constraints. This approach allows to extend the results to measurable payoff functions. In the Balayage approach we consider payoff functions satisfying a growth condition and having at most countably many points of discontinuity.

Finally, we show the existence of an optimal measure μ^*, which is a weighted sum of at most 3 Dirac measures, in the optimization problem over measures in $\mathcal{A}(T)$ under mild conditions on the payoff function. The reduction to discrete measures with at most 3 points of mass simplifies the proof, because we only deal with finite sums instead of integrals with respect to measures $\mu \in \mathcal{A}(T, y)$. We then can identify a consecutive exit time τ^* that embeds μ^* in Y. As a consequence τ^* is optimal in (I.5) and thus, the value of the stopping problem (I.5) is attained in the set of consecutive exit times.

To the best of our knowledge, the idea of using Skorokhod embeddings to analyze optimal stopping problems first appeared in [72], where the authors solve an optimal stopping problem for the geometric Brownian motion, under the Choquet integral, over stopping times that are almost surely finite. When it comes to optimal stopping problems with constraints on the stopping time's distribution, the literature is rather scarce: The seminal book by Shiryaev [63] discusses in Section 4.3 and 4.4 versions of the quickest detection problem with probability constraints. Bayraktar and Miller [6] consider the problem of optimally stopping a Brownian motion with a stopping time whose distribution is atomic with finitely many points of mass. In [8] the authors use optimal transport techniques to treat the problem of optimally stopping a Brownian motion with a stopping time having a fixed specified distribution.

Further stopping problems with an expectation constraint on the stopping time have been solved by Urusov [68]. Let $\theta \in [0,1]$ be the moment at which a standard Brownian motion attains its maximal value on $[0,1]$ and let $\alpha \geq 0$. Then Urusov [68] characterizes the stopping time that minimizes $\mathbb{E}[(\tau-\theta)^+]$ over all stopping times τ satisfying the expectation constraint $\mathbb{E}[(\tau-\theta)^-] \leq \alpha$.

Only few articles deal with reducing the law of the process at the stopping time. In an optimal stopping problem without constraints, where the payoff function is lower semi-continuous and bounded from below, Theorem 2 in [30] allows for a reduction to first exit times of intervals. In this case, the law of the process at the stopping time has at most two mass points. In [21] the authors transform an optimization problem over a class of martingales satisfying a given terminal marginal constraint into an optimal control problem by extending the state space by the conditional distribution process. For the optimal control problem a dynamic programming principle as well as a Hamilton-Jacobi-Bellman equation is obtained. Due to the continuity of the value function in the additional state, the conditional distribution process can be approximated by sums of finitely many Dirac measures. Källblad [36] examines an optimal stopping problem where the distribution of the stopping times is given and the payoff function is path-dependent. Also, by introducing the conditional distribution process she obtains a dynamic programming principle. Under additional regularity assumptions on the payoff function, one can approximate the value function by atomic problems. Henderson, Hobson and Tse [29] consider an agent with prospect theory preferences. The wealth is determined by the value of the Brownian motion W at a stopping time. Under suitable conditions on the so-called value function of the preferences and the probability weighting function, it is optimal to use a stopping time τ such that the law of W_τ is a weighted sum of three Dirac measures.

In Chapter IV we present an extensive example for the results of Chapter II and III. We deal with a sequential testing problem, where an agent has to decide between two simple hypotheses. She collects information until obtaining a significant result and then she accepts the significant hypothesis. In many cases, the observable information until a fixed time t is not significant enough to decide for one hypothesis. Consequently, the decision maker either has to continue collecting information after time t and to hope for a significant result within a reasonable time horizon or she has to end the observation without any result. In particular, she has to compare additional observation costs and the absence of benefits that are associated to a significant result. If we impose a sharp upper bound $T \in (0,\infty)$ on the time until the observation has to be terminated, then it may happen that the observations propose to accept one hypothesis but the available information is not significant enough. Collecting more information then may lead to a significant result within a small additional time horizon. On the other hand, if the observations are significant at time $t \in [0,T)$, the agent can accept H_0 or H_1 before time T and the time horizon is not fully exploited. Hence, we impose a constraint on the average time until the agent has to stop collecting information and thus, respect different scenarios.

We assume that the aggregated information is given by a random walk. The positive and negative increments contribute to H_1 and H_0, respectively. By observing the random walk, the decision maker wants to detect whether the drift is positive or negative and then she accepts H_1 respectively H_0. Once she observes a positive or negative increment, the agent updates her belief about which hypothesis is more likely. Thus, we use the continuous time sequential testing model from Chapter VI.21 in [55] which allows to update the beliefs. More precisely, in the sequential testing model the decision maker continuously observes a Brownian motion X having either drift 0 or κ, $\kappa \neq 0$ and she wants to detect the value of the drift rate b. At the beginning of the observation the agent has a priori beliefs $y \in (0,1)$ and $1-y$ for the hypothesis H_1 and H_0, respectively. Then the a posteriori probability process $Y_t := \mathbb{P}[b=\kappa \,|\, \mathcal{F}_t^X]$, $t \in [0,\infty)$, satisfies $Y_0 = y$ and describes how likely a drift rate of κ is at time t, given all information on the Brownian motion X until time t. Here $(\mathcal{F}_t^X)$ denotes the filtration generated by X.

We impose a threshold $\alpha \in \left(0,\frac{1}{2}\right)$ for the a posteriori probability process Y, which allows the agent to accept H_0 and H_1 only if the hypotheses are significant enough at time t, i.e. if $Y_t \leq \alpha$ or

$Y_t \geq 1-\alpha$. Moreover, we assume that the decision maker gains $\beta \geq 1$ and 1 if she accepts H_1 and H_0, respectively. If she stops the observation process without accepting any hypothesis, she obtains nothing. Therefore, the agent deals with the optimal stopping problem

$$V(T,y) = \sup\left\{\mathbb{E}\big[f(Y_\tau)\big] : \tau \in \mathcal{S}(T),\ Y_0 = y\right\}, \tag{I.6}$$

where $f(x) = \mathbb{1}_{(0,\alpha]}(x) + \beta\mathbb{1}_{[1-\alpha,1)}(x)$ and $\mathcal{S}(T)$ denotes the set of all stopping times with $\mathbb{E}[\tau] \leq T$.

Since the process Y is a solution to a time-homogeneous SDE driven by a Brownian motion $\widetilde{W}$, cf. Chapter VI.21 in [55], we use the results from Chapter II and III to analyze the optimal stopping problem (I.6). We conclude that it is sufficient for the agent to focus on stopping times τ such that the law of Y_τ is a weighted sum of at most three Dirac measures. We show that for some $T \in (0,\infty)$ the value $V(T,y)$ is only attained by a consecutive exit time τ^* and not by first exit times of intervals. In this case, it turns out that the law of Y_{τ^*} has mass points $\alpha, 1-\alpha$ and $b^* \in \left(0, \frac{1}{2}\right]$, which do not depend on the initial value of Y nor on T. Hence, the same three points suffice. If the supremum in (I.6) is attained by a first exit time, then exactly one endpoint of the interval is independent of y and T. If stopping at 3 points entails a higher payoff than using first exit times, then allowing for three outcomes – accepting H_0 or H_1 or terminating the observation with *no* result – increases the expected payoff of the agent.

To identify optimal stopping times and the value function $V(T,y)$, we combine the characterization of (I.6) as a measure optimization problem as well as an optimal stopping problem. This enables us to simplify and shorten the proofs.

In addition, we show that the value function of the optimal stopping problem (I.6) is a solution to the PDE we derive in Chapter II.

Shiryaev considers this sequential testing model in Chapter 4.2 of [63] (see also Chapter VI.21 in [55]) but instead of maximizing a payoff function depending on the a posteriori probability process, he chooses a stopping time and a decision rule in order to minimize the sum of the expected time until the agent stops the observation and the error probabilities of wrong decisions. Here, the optimal stopping time is a first exit time of an interval. Peskir and Gapeev [27] deal with sequentially testing two simple hypotheses about the drift of a Brownian motion on a finite time interval and minimize the same functional as Shiryaev. In this problem, the optimal stopping time is a first exit time of a time-dependent interval.

In the optimal stopping problem (I.5) we can confine to stopping times τ such that the law of Y_τ is a weighted sum of at most 3 Dirac measures. Naturally, the question arises which conditions guarantee that the supremum is attained in the set of first exit times, i.e. stopping at two points is enough. In Chapter V we focus on processes Y with state space $[a,b]$, $-\infty < a < b < \infty$, and derive a sufficient condition for the payoff function f such that a restriction to first exit times in the stopping problem (I.5) is possible. We use the reformulation of this stopping problem as a measure optimization over the set $\mathcal{A}(T)$. If the state space is bounded, then $\mathcal{A}(T) = \mathcal{A}(T,y)$ is the set of all probability measures μ satisfying $\int_a^b x\,\mu(dx) = y$ and $\int_a^b q_y(x)\mu(dx) \leq T$, where y is the initial value of the process Y and q_y depends on y and the speed measure of Y. To deal with maximization problems over measures satisfying integrability constraints, we introduce the concept of Tchebycheff systems, see [38].

Let $u_0,...,u_n\colon [a,b] \to \mathbb{R}$ be continuous. The functions $u_0,...,u_n$ form a *Tchebycheff system* on $[a,b]$ if

$$\det\begin{pmatrix} u_0(x_0) & u_0(x_1) & ... & u_0(x_n) \\ u_1(x_0) & u_1(x_1) & ... & u_1(x_n) \\ \vdots & \vdots & & \vdots \\ u_n(x_0) & u_n(x_1) & ... & u_n(x_n) \end{pmatrix} \tag{I.7}$$

is either strictly positive for all $a \leq x_0 < x_1 < ... < x_n \leq b$ or strictly negative. The functions u_i, $0 \leq i \leq n$, serve as constraint functions for a measure optimization. Denote by $\mathscr{S}$ the set of all non-decreasing, right-continuous functions of bounded total variation. Then $\varsigma \in \mathscr{S}$ is the Stieltjes measure function of a finite measure ν on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In general, ν is not a probability measure. For a continuous function g and $c_0, ..., c_n \in \mathbb{R}$, $0 \le i \le n$, we consider the following problem

$$\text{maximize } \int_a^b g(x) d\varsigma(x) \quad \text{subject to } \varsigma \in \mathscr{S} \text{ and } \int_a^b u_i(x) d\varsigma(x) = c_i,\ 0 \le i \le n. \tag{I.8}$$

Firstly, we study the space $\mathscr{M}_{n+1} \subseteq \mathbb{R}^{n+1}$ of u-moments $\left(\int_a^b u_i(x) d\varsigma(x)\right)_{0 \le i \le n}$, $\varsigma \in \mathscr{S}$, and characterize its boundary and interior points. If both $(u_0, ..., u_n)$ and the extended system $(u_0, ..., u_n, g)$ are Tchebycheff systems, then for any interior point $\mathbf{c} = (c_0, ..., c_n) \in \mathscr{M}_{n+1}$ we can identify the unique maximizer $\varsigma^* \in \mathscr{S}$ in (I.8). The support $\operatorname{supp}(\varsigma^*)$ of ς^* is discrete and if n is even, $\operatorname{supp}(\varsigma^*)$ contains $\frac{n}{2}$ points from (a, b) and exactly one of the endpoints a or b. For odd n the support of ς^* is given by $\frac{n+1}{2}$ points from (a, b) or $\frac{n-1}{2}$ points from (a, b) and both endpoints a and b, depending on the sign of the determinant (I.7). In Section V.1 we present the results from Chapter I and II of [38] which are used to derive ς^*. Furthermore, we extend the proofs from [38] and give more details, because to the best of our knowledge the theory of Tchebycheff systems is not very well-established in the literature on optimal stopping.

We apply the results for Tchebycheff systems to the measure optimization problem over $\mathcal{A}(T)$. The constraints describing $\mathcal{A}(T, y)$ are given by $\int_{\mathbb{R}} u_i(x) \mu(dx) = c_i$, $i = 0, 1$, and $\int_a^b u_2(x) \mu(dx) \le T$, where $u_0(x) = 1, u_1(x) = x$, $u_2(x) = q_y(x)$, $x \in [a, b]$, and $c_0 = 1, c_1 = y$. Here we impose the constraint $\mu([a, b]) = 1$ to guarantee that μ is indeed a probability measures. We show that (u_0, u_1, u_2) constitutes a Tchebycheff system. Moreover, if the payoff function f and the function q_y are smooth enough, then the condition that $\frac{f''}{q_y''}$ is injective on (a, b) is both necessary and sufficient for (u_0, u_1, u_2, f) to be a Tchebycheff system. Moreover, we prove that the unique optimal measure μ^S in $\mathcal{A}(T, y)$ for the optimization problem with additional constraint $\int_{\mathbb{R}} q_y(x) \mu(dx) = S$, $S \in [0, T]$, $\mu \in \mathcal{A}(T, y)$, is a weighted sum of two Dirac measures. One mass point is either given by a or b. This allows to restrict the stopping times in (I.5) to first exit times of intervals if the state space J of Y is compact.

If J is not compact, we approximate the state space by a sequence of compact intervals $J_k \subseteq J$, $k \in \mathbb{N}$. Since we can restrict to measures $\mu \in \mathcal{A}(T)$ that are weighted sums of 3 Dirac measures, there exists $\ell = \ell(\mu) \in \mathbb{N}$ such that the of support of μ is contained in J_ℓ. If the state space is bounded, then the set $\mathcal{A}(T, y)$ only contains measures μ that are centered around y, i.e. $\int_{\mathbb{R}} x\, \mu(dx) = y$. If J is unbounded, then in general the measures in $\mathcal{A}(T, y)$ are not centered around y. We show that on each J_k a reduction to first exit times in (I.5) is possible if (u_0, u_1, u_2, f) is a Tchebycheff system over J_k. This also holds true if the measures are not centered around y. Therefore, if the extended system (u_0, u_1, u_2, f) is a Tchebycheff system over every J_k, $k \in \mathbb{N}$, we can reduce the set of optimal stopping times in (I.5) to first exit times.

Tchebycheff systems are used in different mathematical areas such as approximation theory, e.g. for interpolation methods or quadrature formulas (Section 9, Chapter XI in [38]), and in the theory of inequalities such as generalized Tchebycheff inequalities (cf. Chapter XII–XIV in [38] and the references therein). In [1] and [38, Chapter IX] a continuous function g on a compact interval is approximated in the supremum norm by so-called u-polynomials $\sum_{i=0}^n a_i u_i$, $a_0, ..., a_n \in \mathbb{R}$ for some given functions $u_0, ... u_n$. The best approximation is unique if and only if $(u_0, ..., u_n)$ is a Tchebycheff system. Schumaker [61, Chapter 9] investigates the approximation of smooth functions by Tchebycheffian splines, which are a generalization of polynomial splines.

The parts of this introduction dealing with Chapter II and III are based on [4] and [3], respectively.

II. Optimal Stopping Problems with Expectation Cost Constraints

In many optimal stopping problems the stopping time is restricted to a bounded interval $[0,T]$, $T \in (0,\infty)$, or the stopping problem has infinite time horizon. Whenever a stopping rule is applied repeatedly and independently of the previous stopping times, an average constraint $\mathbb{E}[\tau] \leq T$ on the stopping time τ seems to be more appropriate than a sharp constraint $\tau \leq T$, a.s. For example think of the question of when to stop searching for a parking space. If you face this question whenever you are driving to your work, it is more likely that you impose an average constraint on your searching time than just a sharp upper bound.

In this chapter the process X to stop is an $\mathbb{R}^n$-valued solution to a time-homogeneous stochastic differential equation driven by a Brownian motion, where the drift and the diffusion coefficient are Lipschitz continuous. Here we deal not only with expectation constraints on the stopping time τ but with more general cost constraints of the form $\mathbb{E}\left[\int_0^\tau h(X_s)ds\right] \leq T$ for measurable functions $h\colon \mathbb{R}^n \to (0,\infty)$. Since there is no simple dependence of the constraint on time, we turn the expectation constraint into a path-dependent constraint by extending the state space with the conditional expectation process of the constraint. Consequently, the expectation constraint is transformed into an initial condition. We can interpret the new state variable as a martingale with controlled diffusion coefficient and establish a one-to-one correspondence between the set of stopping times satisfying $\mathbb{E}\left[\int_0^\tau h(X_s)ds\right] \leq T$ and a class of controlled martingales in Section II.2. The correspondence allows us to transform the constrained stopping problem to an unconstrained optimal control problem with extended state space. For the control problem we can formulate a dynamic programming principle (DPP) (see Section II.3) and thus characterize the value function as a solution of the dynamic programming equation (DPE), see Section II.3.3. Moreover, in Section II.4 we state in a classical verification theorem sufficient conditions that guarantee that a solution of the DPE coincides with the value function. In addition, we provide several example for the verification theorem. Using the dynamic programming equation and the verification theorem we construct two families of optimal stopping problems for various constraint functions in Section II.5. In Section II.6 we briefly compare the method of extending the state space with the Lagrange approach. Finally, in Section II.7 we discuss several examples illustrating the scope of our results.

This chapter is based on [4]. Section II.5 is not part of the article [4].

II.1 Optimal Stopping with Expectation Cost Constraints

Let $(W_t)_{t\in[0,\infty)}$ be a d-dimensional Brownian motion on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Denote by $(\mathcal{F}_t)_{t\in[0,\infty)}$ its augmented natural filtration and let $\mathcal{F}_\infty = \bigcup_{t\in[0,\infty)} \mathcal{F}_t$. Let $b\colon \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma\colon \mathbb{R}^n \to \mathbb{R}^{n\times d}$ be Lipschitz-continuous functions and assume that for every $x \in \mathbb{R}^n$ the matrix $(\sigma\sigma^\top)(x) \in \mathbb{R}^{n\times n}$ is positive definite. Here $\sigma^\top(x)$ denotes the transpose of the matrix $\sigma(x)$, $x \in \mathbb{R}^n$. Then there exists a unique $\mathbb{R}^n$-valued strong solution $(X_t^x)_{t\in[0,\infty)}$ of the stochastic differential equation (SDE)

$$dX_t^x = b(X_t^x)dt + \sigma(X_t^x)\cdot dW_t, \qquad X_0^x = x \tag{II.1.1}$$

for every $x \in \mathbb{R}^n$.

Let $h\colon \mathbb{R}^n \to (0,\infty)$ be Borel-measurable and define the process $(H_t^x)_{t\in[0,\infty)}$ by

$$H_t^x = \int_0^t h(X_s^x)\,ds$$

for $x \in \mathbb{R}^n$ fixed. Denote by $\mathcal{T}(T) = \mathcal{T}(T,x)$ and $\mathcal{S}(T) = \mathcal{S}(T,x)$ the set of all $(\mathcal{F}_t)$-stopping times τ with $\mathbb{E}[H_\tau^x] = T$ and $\mathbb{E}[H_\tau^x] \le T$ respectively. In the following we sometimes refer to $\mathcal{T}(T)$ as the set of admissible stopping times.

Assumption. Throughout this chapter we assume that for all $x \in \mathbb{R}^n$ and $t \in [0,\infty)$

$$H_t^x < \infty \quad \text{and} \quad \lim_{s\to\infty} H_s^x = \infty,\ \mathbb{P} - \text{a.s.} \tag{$\mathbf{A}$}$$

Note that Assumption ($\mathbf{A}$) guarantees that $\mathcal{T}(T)$ is non-empty; e.g. the stopping time $\tau = \inf\{s \in [0,\infty)\colon H_s^x > T\}$ satisfies $H_\tau^x = T$, $\mathbb{P}$-a.s., and hence lies in $\mathcal{T}(T)$. If h is bounded and bounded away from zero, i.e. $h\colon \mathbb{R}^n \to [\delta, C]$ with $0 < \delta < C$, then Assumption ($\mathbf{A}$) is satisfied, because $\delta t \le H_t^x \le Ct$ for $t \in [0,\infty)$.

In order to simplify notation, in the following we often write X_t and H_t instead of X_t^x and H_t^x, respectively.

For a Borel-measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$ we consider the following optimal stopping problem with constraint function h

$$V(T,x) = \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}[f(X_\tau^x)], \tag{II.1.2}$$

where $T \in [0,\infty)$ and $x \in \mathbb{R}^n$. Here we use the convention that $\mathbb{E}[f(X_\tau^x)] = -\infty$ if both the negative and the positive part of $f(X_\tau^x)$ have infinite expectation. Notice that by Assumption ($\mathbf{A}$) every $\tau \in \mathcal{S}(T)$ is finite $\mathbb{P}$-a.s. and hence X_τ^x is well-defined. If $\mathbb{P}[H_\infty^x < \infty] > 0$, then it can happen that X_τ^x does not exist, as the following example shows.

Example II.1.1. Let $X_t = x_0 + W_t, t \in [0,\infty)$, be a Brownian motion in $\mathbb{R}^3$ starting in $x_0 = (1,0,0)$. For the constraint function $h\colon \mathbb{R}^3 \to (0,\infty)$, $h(y) = e^{-|y|^2}$ we have that H_∞ is finite $\mathbb{P}$-a.s., because

$$\mathbb{E}[H_\infty] = \mathbb{E}\left[\int_0^\infty h(X_s)ds\right] = \int_0^\infty \frac{1}{(2s+1)^{\frac{3}{2}}} e^{-\frac{1}{2s+1}}\,ds = \int_0^1 e^{-s^2}\,ds \le 1.$$

Now consider the stopping time

$$\tau = \inf\left\{t \in [0,\infty)\colon\ |X_t| = \frac{1}{2}\right\},$$

which satisfies $\mathbb{P}[\tau < \infty] = \frac{1}{2}$ (see [49], Example 7.4.2). Moreover, $\tau \in \mathcal{S}(T)$ for $T \in [1,\infty)$, but $\lim_{t\to\infty} X_t$ does not exist.

Remark II.1.2. For a one-dimensional Brownian motion $(X_t^x)_{t\in[0,\infty)}$ the 0-1-law of Engelbert and Schmidt (see e.g. [26]) implies that if h is locally integrable on $\mathbb{R}$, we have $H_t^x < \infty$, $\mathbb{P}$-a.s., for all $t \in (0,\infty)$ and $x \in \mathbb{R}$, and hence, the first part of Assumption ($\mathbf{A}$) is satisfied.

It turns out to be useful to study also the stopping problem with the equality constraint $\mathbb{E}[H_\tau^x] = T$. We therefore introduce

$$U(T,x) = \sup_{\tau\in\mathcal{T}(T)} \mathbb{E}[f(X_\tau^x)]. \tag{II.1.3}$$

Then $V(T,x) = \sup_{S\le T} U(S,x)$ and the value function V is completely determined by U.

Lemma II.1.3. *Let $x \in \mathbb{R}^n$. Suppose that $T \mapsto U(T,x)$ is concave and $U(T,x) \in \mathbb{R}$ for all $T \in [0,\infty)$. Then*

$$V(T,x) = U\big(T \wedge \widetilde{T}(x), x\big), \quad (T,x) \in [0,\infty) \times \mathbb{R}^n, \tag{II.1.4}$$

where $\widetilde{T}(x) = \inf\big\{t \in [0,\infty) \colon U(t,x) < \sup_{s \le t} U(s,x)\big\}$ and $T \wedge \widetilde{T}(x) = \min\big\{T, \widetilde{T}(x)\big\}$. (We set $\widetilde{T}(x) = \infty$ if the function $t \mapsto U(t,x)$ is non-decreasing, i.e. $U(t,x) \ge U(s,x)$ for all $0 \le s \le t$.)

Proof. Let $x \in \mathbb{R}^n$. The functions $U(.,x)$ and $V(.,x)$ coincide on $\big[0, \widetilde{T}(x)\big)$ by the definition of $\widetilde{T}(x)$. If $\widetilde{T}(x) < \infty$, we have $U(t,x) \le U\big(\widetilde{T}(x), x\big)$ for all $t \ge \widetilde{T}(x)$, because U is concave in T. Observe that by the very definition the function $T \mapsto V(T,x)$ is non-decreasing. Therefore, if $\widetilde{T}(x) < \infty$, then $V(T,x) = V\big(\widetilde{T}(x), x\big)$ for all $T \ge \widetilde{T}(x)$. Since U is concave in T and $U(T,x) \in \mathbb{R}$ for all $T \in [0,\infty)$, U is continuous in T and we have $U\big(\widetilde{T}(x), x\big) = V\big(\widetilde{T}(x), x\big)$. Hence, we obtain (II.1.4). Notice that in this case also $T \mapsto V(T,x)$ is concave. □

In the above derivation of (II.1.4) we have assumed that $T \mapsto U(T,x)$ is concave. Conveniently, concavity turns out to be a consequence of the verification theorem. Notice, however, that one can heuristically show concavity of $T \mapsto U(T,x)$ as follows: let $\tau_1 \in \mathcal{T}(T_1)$ and $\tau_2 \in \mathcal{T}(T_2)$. Flip a coin with probability $\alpha \in (0,1)$ for head, and choose τ_1 if head and τ_2 if tail appears. With the randomized stopping time we can show that $U(\alpha T_1 + (1-\alpha)T_2, x) \ge \alpha U(T_1,x) + (1-\alpha) U(T_2,x)$.

In the next sections we transform problem (II.1.3) into a control problem with an extended state space and derive a dynamic programming equation (DPE) for U. Moreover, we provide a verification theorem that allows to check whether a solution of the DPE coincides with the value function U. The link (II.1.4) allows us then to identify the value function V and to obtain an optimal stopping time for the original problem (II.1.2).

II.2 Every admissible Stopping Time is a first Hitting Time

In this section we establish a one-to-one correspondence between the set of stopping times $\mathcal{T}(m)$ and a class of $(\mathcal{F}_t)$-martingales solving a specific type of SDE with initial value m, where $m \in [0,\infty)$. This correspondence allows us to transform the stopping problems (II.1.2) and (II.1.3).

For every $\tau \in \mathcal{T}(m)$, the process $(M_t)_{t \in [0,\infty]}$ defined by

$$M_t = \mathbb{E}\left[H_\tau \,|\, \mathcal{F}_t\right]$$

is a continuous $(\mathcal{F}_t)$-martingale with $M_\infty = H_\tau$ and $M_0 = \mathbb{E}[H_\tau] = m$. Thus, the martingale representation theorem implies

$$M_t = m + \int_0^t \alpha_s \cdot dW_s,$$

where $(\alpha_t)_{t \in [0,\infty)} = (\alpha_t^1, ..., \alpha_t^d)_{t \in [0,\infty)} \in L^2_{loc}(W)$, i.e. $(\alpha_t)_{t \in [0,\infty)}$ is progressively measurable and there exists a sequence $(\rho_n)_{n \in \mathbb{N}}$ of $(\mathcal{F}_t)$-stopping times with $\rho_n \nearrow \infty$, $\mathbb{P}$-a.s., such that for all $n \in \mathbb{N}$

$$\mathbb{E}\left[\int_0^{\rho_n} |\alpha_s|^2 ds\right] < \infty.$$

Then, the stopping time τ can be characterized as the first time when the process of conditional expectations falls below the process H.

Lemma II.2.1. *Let τ be an $(\mathcal{F}_t)$-stopping time such that H_τ is integrable and let $(M_t)_{t \in [0,\infty)}$ be a continuous version of the process of conditional expectations $\mathbb{E}[H_\tau \,|\, \mathcal{F}_t]$, $t \in [0,\infty)$. Then we have*

$$\tau = \inf\big\{t \in [0,\infty) \colon M_t \le H_t\big\}, \qquad \mathbb{P} - a.s.$$

Proof. Notice that

$$M_t = \mathbb{E}[H_\tau \,|\, \mathcal{F}_t] = \mathbb{1}_{\{\tau \le t\}} H_\tau + \mathbb{1}_{\{\tau > t\}} \mathbb{E}[H_\tau \,|\, \mathcal{F}_t]. \tag{II.2.1}$$

Since $(H_t)_{t\in[0,\infty)}$ is strictly increasing in t we have $H_\tau \le H_t$ on $\{\tau \le t\}$ and $H_t < H_\tau$ on $\{\tau > t\}$. Thus, (II.2.1) implies

$$\begin{aligned}\inf\big\{t\in[0,\infty)\colon M_t \le H_t\big\} &= \inf\big\{q \in \mathbb{Q}\cap[0,\infty)\colon \mathbb{1}_{\{\tau\le q\}}(H_q - H_\tau) + \mathbb{1}_{\{\tau>q\}}\mathbb{E}[H_q - H_\tau \,|\, \mathcal{F}_q] \ge 0\big\}\\ &= \inf\big\{q\in\mathbb{Q}\cap[0,\infty)\colon \tau \le q\big\}\\ &= \tau, \quad \mathbb{P}\text{−a.s.}\end{aligned}$$

□

Notice that $M_t - H_t$ is the expected remaining constraint at time t. Thus, if $M_t = H_t$, then the remaining constraint equals 0 in expectation and thus, we have to stop.

Lemma II.2.1 implies that for $\tau \in \mathcal{T}(m)$ and a continuous version of $M_t := \mathbb{E}[H_\tau \,|\, \mathcal{F}_t]$ the following holds true:

$$\begin{aligned} M_s &> H_s \quad \text{on } \{s<\tau\},\\ M_s &= M_\tau \quad \text{on } \{s \ge \tau\}.\end{aligned}$$

Thus, the process M satisfies

$$dM_t = \mathbb{1}_{\{\tau>t\}}\alpha_t \cdot dW_t = \mathbb{1}_{\{\forall s\le t\colon M_s > H_s\}}\alpha_t\cdot dW_t.$$

Observe that $(M_s)_{s\in[0,\infty)}$ is constant after τ and that $(H_s)_{s\in[0,\infty)}$ is strictly increasing, which implies that $H_s > M_s$ for all $s>\tau$. Hence, $M_t > H_t$ implies $t<\tau$ and as a consequence, $M_s > H_s$ for all $s\le t$. Therefore, $\{\tau > t\} = \{M_t > H_t\}$ and it follows that

$$dM_t = \mathbb{1}_{\{M_t > H_t\}}\alpha_t \cdot dW_t, \quad M_0 = m. \tag{II.2.2}$$

We have thus shown that any stopping time $\tau\in\mathcal{T}(m)$ coincides with the first time a process solving (II.2.2) hits H_t. Indeed, there is a one-to-one correspondence between $(\mathcal{F}_t)$-stopping times with $\mathbb{E}[H_\tau] = m$ and martingales (M_t) satisfying (II.2.2). To establish this correspondence we need the following lemma.

Lemma II.2.2. *Let $(\alpha_t)_{t\in[0,\infty)} = (\alpha_t^1,...,\alpha_t^d)_{t\in[0,\infty)} \in L^2_{loc}(W)$ and $m\in[0,\infty)$. Then there exists a unique strong solution M of* (II.2.2)*. This solution is a non-negative supermartingale.*

Proof. Let $(\alpha_t)_{t\in[0,\infty)} \in L^2_{loc}(W)$, $m\in[0,\infty)$ and $Y_t = m + \int_0^t \alpha_s\cdot dW_s$. Then,

$$\tau = \inf\{t\in[0,\infty)\colon Y_t \le H_t\}$$

defines an $(\mathcal{F}_t)$-stopping time and the stopped process $M_t := Y_{t\wedge\tau}$ satisfies

$$dM_t = \mathbb{1}_{\{\tau>t\}}\alpha_t\cdot dW_t = \mathbb{1}_{\{\forall s\le t\colon Y_s > H_s\}}\alpha_t\cdot dW_t = \mathbb{1}_{\{\forall s\le t\colon M_s = Y_s \text{ and } M_s > H_s\}}\alpha_t \cdot dW_t.$$

As in the derivation of (II.2.2), we obtain that $M_t > H_t$ implies $M_s > H_s$ for all $s\le t$ by the definition of τ. Hence,

$$\{\forall s \le t\colon M_s = Y_s \text{ and } M_s > H_s\} = \{M_t > H_t\}$$

and (M_t) solves (II.2.2).

We next show that M is the unique strong solution of (II.2.2). To this end let $(N_t)_{t\in[0,\infty)}$ be another solution of (II.2.2). Let $\rho = \inf\{t\in[0,\infty)\colon N_t \le H_t\}$. Then $N_{t\wedge\rho} = m + \int_0^{\rho\wedge t}\alpha_s\cdot dW_s$, which implies that $N_t = M_t$ on $[0,\tau\wedge\rho]$. In particular, we have $M_{\tau\wedge\rho} = N_{\tau\wedge\rho} = H_{\tau\wedge\rho}$ and hence $\rho = \tau$. After τ it holds that $N_t = N_\tau = M_\tau = M_t$. Therefore, $N = M$.

Notice that by definition $(M_t)_{t\in[0,\infty)}$ is a continuous, non-negative local martingale. Hence, it is a supermartingale and the limit

$$M_\infty = \lim_{t\to\infty} M_t$$

exists almost surely with $M_\infty \in L^1(\Omega)$. In addition, $(M_t)_{t\in[0,\infty]}$ is a supermartingale and Fatou's lemma implies $\mathbb{E}[M_\infty] \le m$. In particular, almost every path of M is bounded and thus, $\tau := \inf\{t \in [0,\infty) : M_t \le H_t\} < \infty$, $\mathbb{P}$-a.s., by Assumption (**A**). $\square$

Let

$$\mathcal{A} = \big\{\alpha \in L^2_{loc}(W) : \mathbb{E}[H_\tau] = M_0, \text{ where } M \text{ solves (II.2.2) for } \alpha \text{ and } \tau = \inf\{t \in [0,\infty) : M_t \le H_t\}\big\}.$$

Lemma II.2.2 implies that for $\alpha \in \mathcal{A}$ the solution (M_t) of (II.2.2) is a true martingale with $M_t \to M_\infty$ in $L^1(\Omega)$ for $t \to \infty$. Moreover, $M_\infty = M_\tau = H_\tau$ by the definition of τ.

On the other hand, if for $\alpha \in L^2_{loc}(W)$ the solution of (II.2.2) is a true martingale with $M_t \to M_\infty$ in $L^1(\Omega)$ for $t \to \infty$, then $\mathbb{E}[H_\tau] = \mathbb{E}[M_\tau] = M_0$.

Observe that $\mathcal{A}$ is non-empty. The following example shows that $\mathcal{A} \ne L^2_{loc}(W)$.

Example II.2.3. Let $d = 1$ and $h(y) = 1$ for all $y \in \mathbb{R}$. Let

$$\alpha_t = -\mathbb{1}_{\{t<1\}} \frac{W_t}{(1-t)^{\frac{3}{2}}} e^{-\frac{W_t^2}{2(1-t)}}$$

and $m = 2$. Then $\rho_n = \inf\{t \in [0,\infty) : |\alpha_t| \ge n\}$ is a localizing sequence for α and thus, $\alpha \in L^2_{loc}(W)$. Moreover, the solution M of (II.2.2) is given by

$$M_t = \begin{cases} 1 + \frac{1}{\sqrt{1-t}} e^{-\frac{W_t^2}{2(1-t)}}, & t < 1, \\ 1, & t \ge 1. \end{cases}$$

Then $M_t \ge 1$ for all $t \in [0,\infty)$ and $M_1 = 1 = H_1$. Thus, $\tau := \inf\{t \in [0,\infty) : M_t \le t\} = 1$, $\mathbb{P}$-a.s. Moreover, (M_t) is a local martingale, but not a true martingale, because $M_0 = 2$ and $M_1 = 1$, $\mathbb{P}$-a.s. In particular, we have $\mathbb{E}[H_\tau] = \mathbb{E}[\tau] = \mathbb{E}[M_\tau] = 1 \ne 2 = M_0$.

Let $\mathcal{M}(m)$ be the set of all solutions M of (II.2.2) with $(\alpha_t)_{t\in[0,\infty)} \in \mathcal{A}$. The results obtained so far show that one can identify $\mathcal{T}(m)$ with $\mathcal{M}(m)$.

Proposition II.2.4. *There is a one-to-one correspondence between $\mathcal{T}(m)$ and $\mathcal{M}(m)$ given by*

$$M_t = \mathbb{E}[H_\tau \,|\, \mathcal{F}_t], \quad t \in [0,\infty), \qquad \text{and} \qquad \tau = \inf\{t \in [0,\infty) : M_t \le H_t\},$$

where $\tau \in \mathcal{T}(m)$ and $M \in \mathcal{M}(m)$. Moreover, $H_\tau = M_\tau = M_\infty$.

Proof. The statements follow from Lemma II.2.1 and II.2.2, and the discussion preceding and succeeding Lemma II.2.2. $\square$

Remark II.2.5. To emphasize the dependence on α and m, in the following we often write $M^{\alpha,m}$ instead of M.

Remark II.2.6. A version of this one-to-one correspondence is also established in [43, Lemma 1] in the case where $n = d = 1$, $h(y) = 1$ for all $y \in \mathbb{R}$ and where the sets $\mathcal{T}(m)$ and $\mathcal{M}(m)$ are restricted to square integrable stopping times and square integrable progressively measurable processes, respectively.

Example II.2.7. Let $(W_t)_{t\in[0,\infty)}$ be a Brownian motion in $\mathbb{R}^d$. For $x \in \mathbb{R}^d$ let $X_t^x = x + W_t$, $t \in [0,\infty)$. Suppose that the constraint function h is given by $h(y) = 1, y \in \mathbb{R}^d$, and hence $H_t = t$. Denote by ρ_R the first exit time of the ball around 0 with radius $R > |x|$, i.e. $\rho_R = \inf\{t \in [0,\infty) \colon |X_t^x| \geq R\}$. Then, the expected value of ρ_R is given by

$$\mathbb{E}[\rho_R] = \frac{R^2 - |x|^2}{d},$$

see Chapter 4.2.E in [37]. Hence, on $\{\rho_R > t\}$ the process of conditional expectations M_t is given by

$$M_t = \mathbb{E}[\rho_R \,|\, \mathcal{F}_t] = t + \frac{R^2 - |X_t^x|^2}{d} = \frac{R^2 - |x|^2}{d} + \frac{1}{d}\int_0^t -2(X_s^x)^\top \cdot dW_s.$$

Thus, $\alpha_s = -2X_s^x/d$, $s \in [0,\infty)$, and we conclude that $\rho_R = \inf\left\{t \in [0,\infty) \colon M_t^{\alpha,(R^2-|x|^2)/d} \leq t\right\}$ from Proposition II.2.4. In particular, since $(M_t)_{t\in[0,\infty]}$ is a martingale with $\mathbb{E}[H_\tau] = \mathbb{E}[M_\tau] = M_0$, it follows that $\alpha \in \mathcal{A}$.

In dimension one we can extend the above example to exit times of intervals (a,b), $a < x < b$, instead of intervals $(-R,R)$ with $R > |x|$.

Example II.2.8. For $x \in \mathbb{R}$ let $X_t^x = x + W_t$, $t \in [0,\infty)$, where $(W_t)_{t\in[0,\infty)}$ is a one-dimensional Brownian motion. Again let $h(y) = 1, y \in \mathbb{R}$, and, thus, $H_t = t$. The first exit time $\rho(a,b)$ of an interval (a,b), $a < x < b$, has expectation $(b-x)(x-a)$. The associated process of conditional expectations M_t on $\{M_t > t\}$ is given by

$$M_t = t + (b - X_t^x)(X_t^x - a) = (b-x)(x-a) + \int_0^t \alpha_s \, dW_s$$

with $\alpha_s = -2X_s^x + a + b$, $s \in [0,\infty)$, and $(\alpha_s)_{s\in[0,\infty)} \in \mathcal{A}$.

The next example shows that the one-to-one-correspondence stated in Proposition II.2.4 does not hold in general if the constraint function h is not strictly positive, i.e. the cost process $(H_t)_{t\in[0,\infty)}$ is not strictly increasing

Example II.2.9. Let $(W_t)_{t\in[0,\infty)}$ be a one-dimensional Brownian motion starting in 0, let

$$h(y) = \mathbb{1}_{\{|y|>1\}} \qquad \text{and} \qquad \rho = \inf\{t \in [0,\infty) \colon W_t \notin (-1,1)\}.$$

Then we have

$$H_\rho = \int_0^\rho \mathbb{1}_{\{|W_s|>1\}} ds = 0$$

and thus for all $t \in [0,\infty)$

$$M_t = \mathbb{E}[H_\rho \,|\, \mathcal{F}_t] = 0.$$

Hence,

$$\inf\{t \in [0,\infty) \colon M_t \leq H_t\} = 0 < \rho, \qquad \mathbb{P}-\text{a.s.}$$

The Stopping Problem as an Optimal Control Problem

The one-to-one correspondence of Proposition II.2.4 allows to reformulate the optimal stopping problems (II.1.2) and (II.1.3) as optimal control problems. More precisely, we have

$$V(T,x) = \sup\left\{\mathbb{E}\left[f\left(X^x_{\tau^{\alpha,m}}\right)\right] : m \in [0,T],\, \alpha \in \mathcal{A}\right\} \tag{$\mathcal{V}$}$$

and

$$U(T,x) = \sup\left\{\mathbb{E}\left[f\left(X^x_{\tau^{\alpha,T}}\right)\right] : \alpha \in \mathcal{A}\right\}, \qquad (\mathcal{U})$$

where

$$X^x_t = x + \int_0^t b(X^x_s)ds + \int_0^t \sigma(X^x_s)\cdot dW_s,$$

$$M^{\alpha,m}_t = m + \int_0^t \mathbb{1}_{\{M^{\alpha,m}_s > H^x_s\}}\alpha_s \cdot dW_s,$$

$$\tau^{\alpha,m} = \inf\left\{t\in[0,\infty)\colon M^{\alpha,m}_t \le H^x_t\right\}.$$

As we show in the next section, the reformulations $(\mathcal{V})$ and $(\mathcal{U})$ have the advantage, compared to the original versions (II.1.2) and (II.1.3), that they allow to write down a dynamic programming principle.

Remark II.2.10. The idea to make stochastic control problems with expectation constraints accessible for dynamic programming techniques by extending the state space by a conditional expectation process can already be found in the literature. For example, Bouchard, Elie and Touzi [13] (see also [12]) analyze a control problem where a stochastic target has to be attained with prescribed probability. The authors introduce a controlled conditional probability process as a further state variable (playing the role of M here) which allows them to apply the geometric dynamic programming approach of [66].

II.3 Derivation of a Dynamic Programming Equation

The aim of this section is to derive a dynamic programming equation (DPE) for U. For the derivation we need that the value function U is finite and that it satisfies a dynamic programming principle (DPP). We start, therefore, with a subsection providing sufficient conditions for finiteness.

II.3.1 Finiteness of the Value Functions

Note that if the payoff function f is bounded, then the value functions are also bounded. We next give a more general condition, in the one-dimensional case $d = n = 1$, guaranteeing that the value functions V and U are finite.

Let $d = n = 1$ and denote by J the state space of X and by (l, r), $-\infty \le l < r \le \infty$, the interior of J. By assumption we have $\sigma^2(x) > 0$ for all $x \in (l,r)$. Furthermore, we assume that $(1 + |b(x)|)/\sigma^2(x)$ is locally integrable on (l,r) (see conditions (ND)$'$ and (LI)$'$ in Section 5.5.C of [37]). Let X^x be a solution of (II.1.1) with $X^x_0 = x \in (l,r)$ and define the scale function s_x by

$$s_x(y) = \int_x^y \exp\left(-\int_x^z \frac{2b(w)}{\sigma^2(w)}dw\right)dz, \qquad y\in J.$$

Then the process $Z_t := s_x(X^x_t)$, $t\in[0,\infty)$, is a local martingale with state space $s_x(J)$ and

$$dZ_t = \eta(Z_t)\,dW_t, \quad Z_0 = 0,$$

where $\eta = (s'_x\sigma)\circ s_x^{-1}$. Hence, we can convert the optimal stopping problem with reward function f and constraint function h for the process X into an optimal stopping problem with reward function $f\circ s_x^{-1}$ and constraint function $h\circ s_x^{-1}$ for Z. Let

$$q_x(y) = \int_0^y\int_0^z \frac{2\,h\left(s_x^{-1}(w)\right)}{\eta^2(w)}dw\,dz, \quad y \in s_x(J). \qquad \text{(II.3.1)}$$

In the following we show that if $f\circ s_x^{-1}$ is bounded from above by q_x, $x\in(l,r)$, then $V(T,x)<\infty$ for all $T\in[0,\infty)$. More precisely,

Proposition II.3.1. *Let $x \in (l,r)$. If there exists $C \in (0,\infty)$ such that $f\big(s_x^{-1}(y)\big) \le C\big(1+q_x(y)\big)$ for all $y \in s_x(J)$, then $U(T,x) \le V(T,x) < \infty$. If, in addition, $\big|\mathbb{E}\big[f\big(X^x_{\tau^{0,T}}\big)\big]\big| < \infty$, where $\tau^{0,T} = \inf\{t \in [0,\infty)\colon H_t^x \ge T\}$, then $U(T,x) > -\infty$.*

Proof. First notice that the definition of V implies $V(T,x) \ge f(x)$ for all $x \in (l,r)$.

Observe that $q_x \in \mathcal{C}^1\big((l,r)\big)$ and that q_x' is absolutely continuous with weak derivative $2(h \circ s_x^{-1})/\eta^2 > 0$. Hence, we can apply Itô's formula for $\mathcal{C}^1$-functions with absolutely continuous derivatives, cf. Chapter 3, Exercise 7.3 in [37], to show that $q_x(Z_t) - H_t^x$, $t \in [0,\infty)$, is a local martingale. Moreover, we have $\mathbb{E}[q_x(Z_\tau)] \le T$ for all $\tau \in \mathcal{S}(T)$: Let $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for Z. Then Fatou's lemma and the monotone convergence theorem imply that

$$\begin{aligned}\mathbb{E}[q_x(Z_\tau)] = \mathbb{E}\big[\liminf_{n\to\infty} q_x(Z_{\tau\wedge\tau_n\wedge n})\big] &\le \liminf_{n\to\infty} \mathbb{E}[q_x(Z_{\tau\wedge\tau_n\wedge n})] \\ &= \liminf_{n\to\infty} \mathbb{E}[H^x_{\tau\wedge\tau_n\wedge n}] = \mathbb{E}[H^x_\tau] \le T.\end{aligned}$$

Therefore, we have for $T \in [0,\infty)$

$$\begin{aligned}U(T,x) \le V(T,x) = \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}[f(X^x_\tau)] &= \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}\big[f\big(s_x^{-1}(Z_\tau)\big)\big] \\ &\le C\Big(1 + \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}[q_x(Z_\tau)]\Big) \le C(1+T).\end{aligned}$$

Let $\tau^{0,T} = \inf\big\{t \in [0,\infty)\colon H_t^x \ge T\big\}$ and recall that $\tau^{0,T} \in \mathcal{T}(T)$. If $\mathbb{E}\big[f\big(X^x_{\tau^{0,T}}\big)\big] > -\infty$, then we conclude that $U(T,x) \ge \mathbb{E}\big[f\big(X^x_{\tau^{0,T}}\big)\big] > -\infty$. □

The following example shows that the condition from Proposition II.3.1 is sharp if X is a Brownian motion.

Example II.3.2. For a one-dimensional Brownian motion W we have $q_0(y) = y^2$. Consider the optimal stopping problem (II.1.2) for $f(y) = |y|^{2+\varepsilon}$, $\varepsilon > 0$, constraint function $h(y) = 1$, i.e. $H_t = t$, and the Brownian motion W. For every $T \in (0,\infty)$ the first time $\rho(a,T)$, $a \in (0,\infty)$, when W hits a or $-\frac{T}{a}$ has expectation T. Hence,

$$V(T,0) \ge U(T,0) \ge \sup_{a\in(0,\infty)} \mathbb{E}\big[f\big(W_{\rho(a,T)}\big)\big] = \sup_{a\in(0,\infty)} \left\{a^{2+\varepsilon}\frac{T}{a^2+T} + \frac{a^2}{a^{2+\varepsilon}}\frac{T^{2+\varepsilon}}{a^2+T}\right\} = \infty.$$

II.3.2 Dynamic Programming Principle for U

For the transformed problem $(\mathcal{U})$ one can formulate the following dynamic programming principle:

DPP for U. Let U be measurable and $|U(T,x)| < \infty$ for all $(T,x) \in [0,\infty) \times \mathbb{R}^n$. We say that U satisfies the DPP if for any family $(\theta^\alpha)_{\alpha\in\mathcal{A}}$ of $(\mathcal{F}_t)$-stopping times we have

$$U(T,x) = \sup_{\alpha\in\mathcal{A}} \mathbb{E}\left[\mathbb{1}_{\{\tau^{\alpha,T}\le\theta^\alpha\}} f\left(X^x_{\tau^{\alpha,T}}\right) + \mathbb{1}_{\{\theta^\alpha<\tau^{\alpha,T}\}} U\left(M^{\alpha,T}_{\theta^\alpha} - H^x_{\theta^\alpha}, X^x_{\theta^\alpha}\right)\right]. \tag{II.3.2}$$

Note that on $\{\theta^\alpha < \tau^{\alpha,T}\}$ it holds that $M^{\alpha,T}_{\theta^\alpha} > H^x_{\theta^\alpha}$.

It is a strong assumption to assume that U satisfies the DPP (II.3.2). We use the DPP in order to derive the dynamic programming equation. However, our main result, the verification theorem presented in Section II.4, does not suppose the DPP to be satisfied.

Remark II.3.3. Inspired by the article [4], which is the basis for this chapter, Bayraktar and Yao [7] show that the value function V of an optimal stopping problem with expectation constraint is continuous if the payoff function f is Lipschitz continuous and, in addition, f and the constraint function h are sufficiently nice. The continuity of V allows them to provide a proof of a dynamic programming principle for the value function V. Moreover, they characterize V as a viscosity supersolution of the associated fully non-linear Hamilton-Jacobi-Bellman (HJB) equation and as a viscosity subsolution to the HJB equation, in which the Hamiltonian is replaced by its upper semi-continuous envelope.

II.3.3 The Dynamic Programming Equation for U

The DPP (II.3.2) allows to derive a dynamic programming equation for U. In order to do so, we denote by $\mathcal{L}$ the generator of the Markov process X, i.e.

$$\mathcal{L}u(x) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 u}{\partial x_i \partial x_j}(x)\sum_{l=1}^{d}\sigma_{il}(x)\sigma_{jl}(x) + \sum_{i=1}^{n} b_i(x)\frac{\partial u}{\partial x_i}(x)$$

for suitable functions $u \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$. For a function $u \in \mathcal{C}^2\big((0,\infty)\times\mathbb{R}^n, \mathbb{R}\big)$ we use the notation

$$\begin{aligned}
u_T(T,x) &= \frac{\partial u}{\partial T}(T,x),\\
u_{TT}(T,x) &= \frac{\partial^2 u}{\partial T^2}(T,x),\\
u_{Tx_i}(T,x) &= \frac{\partial^2 u}{\partial T \partial x_i}(T,x),\\
\nabla_x u(T,x) &= \left(\frac{\partial u}{\partial x_i}(T,x)\right)_{i=1}^{n},\\
\nabla_x u_T(T,x) &= \big(u_{Tx_i}(T,x)\big)_{i=1}^{n}
\end{aligned}$$

and for a matrix $A \in \mathbb{R}^{k\times l}$, $k, l \in \mathbb{N}$, its transpose is denoted by $A^\top$.

Proposition II.3.4. *Assume that h is continuous. If $U \in \mathcal{C}^2\big((0,\infty)\times\mathbb{R}^n\big)$ and U satisfies the DPP* (II.3.2)*, then*

1. *U is a supersolution to*
$$h(x)U_T(T,x) - \mathcal{L}U(T,x) + \frac{\left|\sigma^\top(x)\cdot\nabla_x U_T(T,x)\right|^2}{2U_{TT}(T,x)} = 0 \qquad \text{(II.3.3)}$$
on $(0,\infty)\times\mathbb{R}^n$ with initial condition $U(0,x) = f(x)$. Moreover, U is concave in T and
$$\big\{(T,x)\colon U_{TT}(T,x) = 0\big\} \subseteq \big\{(T,x)\colon \nabla_x U_T(T,x) = 0 \in \mathbb{R}^n\big\}.$$
Here we set $\left|\sigma^\top(x)\cdot\nabla_x U_T(T,x)\right|^2 / U_{TT}(T,x) = 0$ if both the numerator and the denominator equal 0*.*

2. *If, in addition, $\left|\sigma^\top\cdot\nabla_x U_T\right|^2 / U_{TT}$ is continuous on $(0,\infty)\times\mathbb{R}^n$, then U is a solution to* (II.3.3) *on $(0,\infty)\times\big\{x \in \mathbb{R}^n\colon \exists\, T \in (0,\infty)$ such that $(T,x) \in \bar{A}\big\}$, where $\bar{A}$ denotes the closure of $A := \big\{(T,x)\colon U_{TT}(T,x) < 0\big\}$ in $(0,\infty)\times\mathbb{R}^n$.*

Proof. 1. The initial condition is satisfied, because $T = 0$ is equivalent to stopping directly. In order to prove that U is a supersolution of (II.3.3) let $(T,x) \in (0,\infty)\times\mathbb{R}^n$, write X instead of X^x in the following and consider the control $\alpha_s = \mathbb{1}_{\{s\le 1\}}a^\top$, with $a \in \mathbb{R}^d$. Then, $\alpha \in \mathcal{A}$. Let

$$\theta^\alpha = \inf\left\{s \in [0,\infty)\colon |X_s - x| \ge 1 \text{ or } M_s^{\alpha,T} - H_s \notin \left[\frac{T}{2}, 2T\right]\right\}.$$

Since $U \in \mathcal{C}^2\big((0,\infty)\times\mathbb{R}^n\big)$ Itô's formula implies that for $t \in (0,1)$

$$\begin{aligned}
&U\big(M_{t\wedge\theta^\alpha}^{\alpha,T} - H_{t\wedge\theta^\alpha}, X_{t\wedge\theta^\alpha}\big) - U(T,x)\\
&= \int_0^{t\wedge\theta^\alpha} -U_T\big(M_s^{\alpha,T} - H_s, X_s\big)dH_s + \int_0^{t\wedge\theta^\alpha} U_T\big(M_s^{\alpha,T} - H_s, X_s\big)dM_s^{\alpha,T}\\
&\quad + \int_0^{t\wedge\theta^\alpha} \big(\nabla_x U\big(M_s^{\alpha,T} - H_s, X_s\big)\big)^\top \cdot dX_s + \sum_{i=1}^{n}\int_0^{t\wedge\theta^\alpha} \frac{\partial^2 U}{\partial T \partial x_i}\big(M_s^{\alpha,T} - H_s, X_s\big)d\langle M^{\alpha,T}, X^i\rangle_s\\
&\quad + \frac{1}{2}\int_0^{t\wedge\theta^\alpha} U_{TT}\big(M_s^{\alpha,T} - H_s, X_s\big)d\langle M^{\alpha,T}, M^{\alpha,T}\rangle_s + \frac{1}{2}\sum_{i,j=1}^{n}\int_0^{t\wedge\theta^\alpha} \frac{\partial^2 U}{\partial x_i \partial x_j}\big(M_s^{\alpha,T} - H_s, X_s\big)d\langle X^i, X^j\rangle_s
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^{t\wedge\theta^\alpha} \Big(\mathbb{1}_{\{M_s^{\alpha,T} > H_s\}} U_T\big(M_s^{\alpha,T} - H_s, X_s\big)\, a^\top + \big(\nabla_x U\big(M_s^{\alpha,T} - H_s, X_s\big)\big)^\top \cdot \sigma(X_s) \Big) \cdot dW_s \\
&\quad + \int_0^{t\wedge\theta^\alpha} \Big(-h\,(X_s)\, U_T + \mathcal{L}U + \mathbb{1}_{\{M_s^{\alpha,T} > H_s\}} \Big(\frac{|a|^2}{2} U_{TT} + (\nabla_x U_T)^\top \cdot \sigma(X_s) \cdot a \Big) \Big) \big(M_s^{\alpha,T} - H_s, X_s\big) ds.
\end{aligned}
$$

By the choice of θ^α the indicator functions equal 1. This equality and the DPP for U (II.3.2) yield that

$$
\begin{aligned}
0 &\geq \mathbb{E}\Big[U\big(M_{t\wedge\theta^\alpha}^{\alpha,T} - H_{t\wedge\theta^\alpha}, X_{t\wedge\theta^\alpha}\big) - U(T,x) \Big] \\
&\geq \mathbb{E}\Big[\int_0^{t\wedge\theta^\alpha} \Big(-h\,(X_s^x)\, U_T + \mathcal{L}U + \frac{|a|^2}{2} U_{TT} + (\nabla_x U_T)^\top \cdot \sigma(X_s) \cdot a \Big) \big(M_s^{\alpha,T} - H_s, X_s\big) ds \Big] \\
&\quad + \ \mathbb{E}\Big[\int_0^{t\wedge\theta^\alpha} \Big(U_T\big(M_s^{\alpha,T} - H_s, X_s\big)\, a + \big(\nabla_x U\big(M_s^{\alpha,T} - H_s, X_s\big)\big)^\top \cdot \sigma(X_s) \Big) \cdot dW_s \Big] .
\end{aligned}
\tag{II.3.4}
$$

The stochastic integral has expectation 0, because the integrand is bounded on the stochastic interval $[0, t\wedge\theta^\alpha]$.

By the pathwise continuity of X_s and $M_s^{\alpha,T}$ and the boundedness of the integrand in the Lebesgue integral we obtain, after first dividing by t and then taking the limit $t \downarrow 0$, that

$$
-h(x)U_T(T,x) + \mathcal{L}U(T,x) + \frac{|a|^2}{2} U_{TT}(T,x) + \big(\nabla_x U_T(T,x)\big)^\top \cdot \sigma(x) \cdot a \leq 0
$$

for all $a \in \mathbb{R}^d$. Thus,

$$
-h(x)U_T(T,x) + \mathcal{L}U(T,x) + \sup_{a\in\mathbb{R}^d} \Big\{ \frac{|a|^2}{2} U_{TT}(T,x) + \big(\nabla_x U_T(T,x)\big)^\top \cdot \sigma(x) \cdot a \Big\} \leq 0. \tag{II.3.5}
$$

In particular, the supremum is finite which shows the concavity of U in T and

$$
\begin{aligned}
\big\{(T,x) \colon U_{TT}(T,x) = 0\big\} &\subseteq \Big\{(T,x) \colon \big(\nabla_x U_T(T,x)\big)^\top \cdot \sigma(x) = 0 \in \mathbb{R}^d\Big\} \\
&= \big\{(T,x) \colon \nabla_x U_T(T,x) = 0 \in \mathbb{R}^n\big\}.
\end{aligned}
$$

The last conclusion follows from the fact that $(\sigma\sigma^\top)(x)$ is positive definite. Inequality (II.3.5) simplifies to

$$
-h(x)U_T(T,x) + \mathcal{L}U(T,x) - \frac{\left|\sigma^\top(x) \cdot \nabla_x U_T(T,x)\right|^2}{2U_{TT}(T,x)} \leq 0
$$

if $U_{TT}(T,x) < 0$ and $-h(x)U_T(T,x) + \mathcal{L}U(T,x) \leq 0$ if $U_{TT}(T,x) = 0$. To conclude, U is a supersolution to (II.3.3), if we set $\left|\sigma^\top(x) \cdot \nabla_x U_T(T,x)\right|^2 / U_{TT}(T,x) = 0$ in the case that both expressions are 0.

2. Let $A = \big\{(T,x) \in (0,\infty) \times \mathbb{R}^n \colon U_{TT}(T,x) < 0\big\}$. In order to prove that U is a subsolution to (II.3.3) on $\bar{A}$ if $\left|\sigma^\top \cdot \nabla_x U_T\right|^2 / U_{TT}$ is continuous, we first assume that there exists $(T_0,x_0) \in (0,\infty) \times \mathbb{R}^n$ with $U_{TT}(T_0,x_0) < 0$ and

$$
h(x_0)U_T(T_0,x_0) - \mathcal{L}U(T_0,x_0) + \frac{\left|\sigma^\top(x_0) \cdot \nabla_x U_T(T_0,x_0)\right|^2}{2U_{TT}(T_0,x_0)} > 0. \tag{II.3.6}
$$

Define for $(T,x) \in (0,\infty) \times \mathbb{R}^n$

$$
\varphi(T,x) = U(T,x) + |T - T_0|^4 + |x - x_0|^4.
$$

Then $\varphi \in \mathcal{C}^2\big((0,\infty)\times\mathbb{R}^n\big)$ and (II.3.6) holds also if U is replaced with φ. Moreover, the continuity of the derivatives implies that

$$h(x)\varphi_T(T,x) - \mathcal{L}\varphi(T,x) + \frac{\left|\sigma^\top(x)\cdot\nabla_x\varphi_T(T,x)\right|^2}{2\varphi_{TT}(T,x)} > 0 \tag{II.3.7}$$

and

$$\varphi_{TT}(T,x) < 0 \tag{II.3.8}$$

on $\mathcal{N}_{2r} := (T_0-2r, T_0+2r)\times B_{2r}(x_0) \subseteq (0,\infty)\times\mathbb{R}^n$ for some $r>0$, where $B_{2r}(x_0) = \{z\in\mathbb{R}^n\colon |z-x_0|<2r\}$. Now let $\alpha\in\mathcal{A}$ and $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for α. In the following we write M^α and X instead of M^{α,T_0} and X^{x_0}, respectively. Define the stopping times

$$\begin{aligned}\theta^\alpha &= \inf\big\{t\in[0,\infty)\colon (M^\alpha_t - H_t, X_t)\notin\mathcal{N}_r\big\},\\ \theta^\alpha_n &= \theta^\alpha\wedge\tau_n\wedge n,\quad n\in\mathbb{N}.\end{aligned}$$

Applying Itô's formula to φ leads to

$$\begin{aligned}&U(T_0,x_0) = \varphi(T_0,x_0)\\ &\quad= \mathbb{E}\left[\varphi(M^\alpha_{\theta^\alpha_n} - H_{\theta^\alpha_n}, X_{\theta^\alpha_n})\right]\\ &\qquad+\mathbb{E}\left[\int_0^{\theta^\alpha_n}\left(h(X_s)\,\varphi_T - \mathcal{L}\varphi - \frac{|\alpha_s|^2}{2}\varphi_{TT} - (\nabla_x\varphi_T)^\top\cdot\sigma(X_s)\cdot\alpha_s^\top\right)(M^\alpha_s - H_s, X_s)ds\right]\\ &\quad\geq \mathbb{E}\left[\varphi(M^\alpha_{\theta^\alpha_n} - H_{\theta^\alpha_n}, X_{\theta^\alpha_n})\right]\\ &\qquad+\mathbb{E}\left[\int_0^{\theta^\alpha_n}\left(h(X_s)\,\varphi_T - \mathcal{L}\varphi - \sup_{a\in\mathbb{R}^d}\left\{\frac{|a|^2}{2}\varphi_{TT} + (\nabla_x\varphi_T)^\top\cdot\sigma(X_s)\cdot a\right\}\right)(M^\alpha_s - H_s, X_s)ds\right].\end{aligned} \tag{II.3.9}$$

Here we use that for all $n\in\mathbb{N}$ the stochastic integral

$$N^n_t := \int_0^{t\wedge\theta^\alpha_n}\Big(\varphi_T\alpha_s + (\nabla_x\varphi)^\top\cdot\sigma(X_s)\Big)(M^\alpha_s - H_s, X_s)\cdot dW_s,\quad t\in[0,\infty),$$

is a martingale. Indeed, by Proposition 1.23, Chapter IV in [56] it suffices to show that the quadratic variation $\langle N^n, N^n\rangle$ is integrable, i.e. $\mathbb{E}[\langle N^n,N^n\rangle_\infty] < \infty$. Then, $(N^n_t)_{t\in[0,\infty)}$ is an L^2-bounded martingale and thus, $N^n_t\to N^n_\infty$ in $L^2(\Omega)$ and $\mathbb{P}$-a.s. for $t\to\infty$. Moreover, we have $\mathbb{E}[N^n_\infty] = \mathbb{E}[N^n_0] = 0$. The boundedness of the derivatives on $[0,\theta^\alpha]$ implies

$$\begin{aligned}\mathbb{E}[\langle N^n,N^n\rangle_\infty] &= \mathbb{E}\left[\int_0^{\theta^\alpha_n}\left|\varphi_T\alpha_s + (\nabla_x\varphi)^\top\cdot\sigma(X_s)\right|^2(M^\alpha_s - H_s, X_s)ds\right]\\ &\leq 2\mathbb{E}\left[\int_0^{\theta^\alpha\wedge\tau_n\wedge n}\left(|\alpha_s|^2\varphi_T^2(M^\alpha_s - H_s, X_s) + \left|\big(\nabla_x\varphi(M^\alpha_s - H_s, X_s)\big)^\top\cdot\sigma(X_s)\right|^2\right)ds\right]\\ &\leq 2C\mathbb{E}\left[\int_0^{\tau_n}|\alpha_s|^2ds\right] + 2Cn\\ &< \infty\end{aligned}$$

by the choice of τ_n. By (II.3.8) the supremum in (II.3.9) is given by

$$-\frac{\left|\sigma^\top(X_s)\cdot\nabla_x\varphi_T(M^\alpha_s - H_s, X_s)\right|^2}{2\varphi_{TT}(M^\alpha_s - H_s, X_s)}.$$

Therefore, we conclude from (II.3.7) that

$$U(T_0, x_0) \geq \mathbb{E}\big[\varphi\big(M^\alpha_{\theta^\alpha_n} - H_{\theta^\alpha_n}, X_{\theta^\alpha_n}\big)\big]$$

for all $n \in \mathbb{N}$. Since φ is bounded on $\overline{\mathcal{N}}_r$, taking the limit $n \to \infty$ results in

$$\begin{aligned} U(T_0, x_0) &\geq \mathbb{E}\big[\varphi\big(M^\alpha_{\theta^\alpha} - H_{\theta^\alpha}, X_{\theta^\alpha}\big)\big] \\ &\geq \kappa + \mathbb{E}\big[U\big(M^\alpha_{\theta^\alpha} - H_{\theta^\alpha}, X_{\theta^\alpha}\big)\big], \end{aligned}$$

where $\kappa := \min_{(T,x)\in\partial\mathcal{N}_r}(\varphi - U)(T, x)$. Note that $\kappa > 0$ by the definition of φ. Since α was arbitrary and κ is independent of α, this contradicts the DPP for U. Hence, we have shown that $U(T, x)$ is a subsolution to (II.3.3) on A. The continuity of h, U_T, $\mathcal{L}U$ and $\left|\sigma^\top \cdot \nabla_x U_T\right|^2 / U_{TT}$ implies that U satisfies

$$h(x)U_T(T, x) - \mathcal{L}U(T, x) + \frac{\left|\sigma^\top(x) \cdot \nabla_x U_T(T, x)\right|^2}{2U_{TT}(T, x)} \leq 0$$

for $(T, x) \in \bar{A}$.

Finally, let $(T_0, x_0) \in \big(\bar{A}\big)^c \cap \big((0,\infty) \times \big\{x \in \mathbb{R}^n : \exists T \in (0,\infty) \text{ such that } (T, x) \in \bar{A}\big\}\big)$. Then there exists a neighborhood $\mathcal{N}$ of (T_0, x_0) with $\overline{\mathcal{N}} \subseteq (0,\infty) \times \mathbb{R}^n$ such that $U_{TT} = 0$ on $\overline{\mathcal{N}}$ and $\big((0,\infty) \times \{x_0\}\big) \cap \overline{\mathcal{N}} \cap \bar{A} \neq \emptyset$. In particular, we have $\nabla_x U_T = 0$ on $\overline{\mathcal{N}}$ by the first part of the proof. Hence, the value function U is linear or constant in T and there exist $c \in \mathbb{R}$ and $g \in \mathcal{C}^2(\mathbb{R}^n)$ such that

$$U(T, x) = cT + g(x), \qquad (T, x) \in \overline{\mathcal{N}}.$$

Therefore,

$$h(x_0)U_T(T_0, x_0) - \mathcal{L}U(T_0, x_0) = c\,h(x_0) - \mathcal{L}g(x_0). \tag{II.3.10}$$

On the other hand there exists $T \in (0,\infty)$ such that $(T, x_0) \in \overline{\mathcal{N}} \cap \bar{A}$ with $U(T, x_0) = cT + g(x_0)$ by the continuity of U and

$$h(x_0)U_T(T, x_0) - \mathcal{L}U(T, x_0) = c\,h(x_0) - \mathcal{L}g(x_0) \leq 0 \tag{II.3.11}$$

by the previous part. Combining (II.3.10) and (II.3.11) results in

$$h(x_0)U_T(T_0, x_0) - \mathcal{L}U(T_0, x_0) \leq 0.$$

To sum up, this together with the first part of the Proposition implies that U is a solution to (II.3.3) on $(0,\infty) \times \big\{x \in \mathbb{R}^n : \exists T \in (0,\infty) \text{ such that } (T, x) \in \bar{A}\big\}$. $\square$

Remark II.3.5.

a) In general $U \in \mathcal{C}^2\big((0,\infty) \times \mathbb{R}^n\big)$ is not a solution to (II.3.3) on the set $(0,\infty) \times \{x \in \mathbb{R}^n : (T, x) \notin \bar{A} \text{ for all } T \in (0,\infty)\}$, see Subsection II.7.2 for a counterexample. In Lemma A.1.1 in Appendix A.1 we state sufficient conditions for U being a solution to (II.3.3) on the whole set $(0,\infty) \times \mathbb{R}^n$.

b) Let h be continuous. Then the continuity condition for $\left|\sigma^\top \cdot \nabla_x U_T\right|^2 / U_{TT}$ in the second part of Proposition II.3.4 is necessary: Let $u \in \mathcal{C}^2\big((0,\infty) \times \mathbb{R}^n\big)$ be a solution of (II.3.3) and let $(T, x) \in (0,\infty) \times \mathbb{R}^n$. If $u_{TT}(T, x) \neq 0$, then the quotient is continuous in (T, x). Now assume that $u_{TT}(T, x) = 0$ and observe that

$$\frac{\left|\sigma^\top(y) \cdot \nabla_x u_T(t, y)\right|^2}{u_{TT}(t, y)} = \begin{cases} 2\big(\mathcal{L}u(t, y) - h(y)u_T(t, y)\big), & \text{if } u_{TT}(t, y) \neq 0, \\ 0, & \text{if } u_{TT}(t, y) = 0, \end{cases}$$

for all $(t, y) \in (0,\infty) \times \mathbb{R}^n$.

Thus, for every sequence $(T_n, x_n)_{n\in\mathbb{N}} \subseteq (0,\infty)\times\mathbb{R}^n$ with $(T_n, x_n) \to (T, x)$ as $n\to\infty$ we have

$$\left|\frac{\left|\sigma^\top(x_n)\cdot\nabla_x u_T(T_n,x_n)\right|^2}{u_{TT}(T_n,x_n)}\right| \le 2\left|\mathcal{L}u(T_n,x_n) - h(x_n)u_T(T_n,x_n)\right|$$
$$\xrightarrow[n\to\infty]{} 2\left|\mathcal{L}u(T,x) - h(x)u_T(T,x)\right| = 0,$$

because u solves (II.3.3) and $u_{TT}(T,x) = 0$. Hence,

$$\lim_{n\to\infty}\frac{\left|\sigma^\top(x_n)\cdot\nabla_x u_T(T_n,x_n)\right|^2}{u_{TT}(T_n,x_n)} = 0 = \frac{\left|\sigma^\top(x)\cdot\nabla_x u_T(T,x)\right|^2}{u_{TT}(T,x)},$$

which implies the continuity of $\left|\sigma^\top(x)\cdot\nabla_x u_T(T,x)\right|^2/u_{TT}(T,x)$.
In Example II.7.1 the continuity condition is not satisfied and the value function is only a supersolution to (II.3.3).

Remark II.3.6. The convention that $\left|\sigma^\top(x)\cdot\nabla_x U_T(T,x)\right|^2/U_{TT}(T,x) = 0$ if $U_{TT}(T,x) = 0$ is justified by the proof of the first part of Proposition II.3.4. There it is shown that

$$\sup_{a\in\mathbb{R}^d}\left\{\frac{|a|^2}{2}U_{TT}(T,x) + \left(\nabla_x U_T(T,x)\right)^\top\cdot\sigma(x)\cdot a\right\} = \begin{cases} -\frac{\left|\sigma^\top(x)\cdot\nabla_x U_T(T,x)\right|^2}{2U_{TT}(T,x)}, & \text{if } U_{TT}(T,x) < 0,\\ 0, & \text{if } U_{TT}(T,x) = 0.\end{cases}$$

Remark II.3.7. If one replaces Equation (II.3.3) with

$$h(x)U_T(T,x) - \mathcal{L}U(T,x) - \sup_{a\in\mathbb{R}^d}\left\{\frac{|a|^2}{2}U_{TT}(T,x) + \left(\nabla_x U_T(T,x)\right)^\top\cdot\sigma(x)\cdot a\right\} = 0, \qquad \text{(II.3.12)}$$

then all the statements of Proposition II.3.4 remain true.

Remark II.3.8. The PDE (II.3.12) is also derived in [43, Corollary 3]. In [43, Theorem 4] the value function U is characterized as the unique viscosity solution of (II.3.12) with initial condition $U(0,\cdot) = f$ in the case where $d = n = 1$, $b(x) = 0$, $\sigma(x) = 1$ and $h(x) = 1$ for all $x\in\mathbb{R}$ and where f satisfies a growth condition.

II.4 Verification

In this section we state a classical verification theorem for U and V. Recall Remark II.3.5 a) and Lemma A.1.1 from Appendix A.1: if for all (T,x) there exists an optimal control for problem $(\mathcal{U})$, then U is a solution of (II.3.3) on the whole domain $(0,\infty)\times\mathbb{R}^n$. Therefore, for a verification it is natural to look for a solution of the PDE (II.3.3) on the whole set $(0,\infty)\times\mathbb{R}^n$.

We first show that if $u\in\mathcal{C}^2\big((0,\infty)\times\mathbb{R}^n\big)\cap\mathcal{C}\big([0,\infty)\times\mathbb{R}^n\big)$ is a solution of the PDE (II.3.3), and some additional mild conditions are satisfied, then u coincides with the value function of the optimal control problem $(\mathcal{U})$. The relation (II.1.4) allows us then to identify the value function V.

Theorem II.4.1. *Let $u\in\mathcal{C}^2\big((0,\infty)\times\mathbb{R}^n\big)\cap\mathcal{C}\big([0,\infty)\times\mathbb{R}^n\big)$ be a function that is concave in T, satisfies $u(0,.) = f$ and $\big\{(T,x)\colon u_{TT}(T,x) = 0\big\}\subseteq\big\{(T,x)\colon \nabla_x u_T(T,x) = 0\in\mathbb{R}^n\big\}$. Moreover, assume that u has linear growth in T and polynomial growth in x, i.e. there exist $C\in(0,\infty)$ and $p\ge 1$ such that*

$$|u(T,x)| \le C\big(1 + T + |x|^p\big). \qquad \text{(II.4.1)}$$

Let $(T,x)\in(0,\infty)\times\mathbb{R}^n$. Suppose that for every $\tau\in\mathcal{S}(T)$ the family

$$\mathcal{X}^p_\tau := \big\{\left|X^x_\vartheta\right|^p\colon \vartheta\ (\mathcal{F}_t)\text{-stopping time, } \vartheta\le\tau \text{ a.s.}\big\}$$

is uniformly integrable.

1. *If u is a supersolution of* (II.3.3), *then $u(T,x) \geq U(T,x)$.*

2. *Assume that u is a solution of* (II.3.3).

   a) *If*

$$\alpha_s^* := -\mathbb{1}_{\{u_{TT}<0\}} \left(\frac{(\nabla_x u_T)^\top \cdot \sigma}{u_{TT}} \right) (M_s^* - H_s^x, X_s^x) \in \mathcal{A}, \tag{II.4.2}$$

   where

$$dM_s^* = -\mathbb{1}_{\{M_s^* > H_s\}} \left(\mathbb{1}_{\{u_{TT}<0\}} \frac{(\nabla_x u_T)^\top \cdot \sigma}{u_{TT}} \right) (M_s^* - H_s^x, X_s^x) \cdot dW_s, \quad M_0^* = T, \tag{II.4.3}$$

   then $u(T,x) = U(T,x)$ and (α_s^*) is an optimal control. The corresponding optimal stopping time τ^* in (II.1.3) *is given by*

$$\tau^* = \left\{ t \in [0,\infty) \colon M_t^* \leq H_t^x \right\}.$$

   b) *If* (II.4.2) *holds for every $S \in [0,T]$, where M^* is given by* (II.4.3) *with $M_0^* = S$, then*

$$V(T,x) = \sup_{S\in[0,T]} u(S,x) = u(T \wedge \widetilde{T}(x), x), \tag{II.4.4}$$

   where $\widetilde{T}(x) = \inf\{t \in [0,\infty) \colon u_T(t,x) \leq 0\}$; and an optimal control (α_s^*, m^*) for $V(T,x)$ is given by

$$\alpha_s^* = -\mathbb{1}_{\{u_{TT}<0\}} \left(\frac{(\nabla_x u_T)^\top \cdot \sigma}{u_{TT}} \right) (M_s^* - H_s^x, X_s^x), \qquad m^* = T \wedge \widetilde{T}(x),$$

   where

$$dM_s^* = -\mathbb{1}_{\{M_s^* > H_s\}} \left(\mathbb{1}_{\{u_{TT}<0\}} \frac{(\nabla_x u_T)^\top \cdot \sigma}{u_{TT}} \right) (M_s^* - H_s^x, X_s^x) \cdot dW_s, \quad M_0^* = m^*.$$

   The corresponding optimal stopping time τ^* in (II.1.2) *is given by*

$$\tau^* = \inf \left\{ t \in [0,\infty) \colon M_t^* \leq H_t^x \right\}.$$

Proof. 1. We first show that every concave supersolution u of (II.3.3) with initial condition $u(0,.) = f$ satisfying the assumptions given in the theorem dominates U. Let $(T,x) \in (0,\infty) \times \mathbb{R}^n$. In the following we write H and X instead of H^x and X^x. For $(\alpha_s)_{s\in[0,\infty)} \in \mathcal{A}$ let $(M_s^{\alpha,T})_{s\in[0,\infty)}$ be the unique strong solution of (II.2.2) and let $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for α. For every $n \in \mathbb{N}$ define the stopping times

$$\rho_n = \inf \left\{ t \in [0,\infty) \colon M_t^{\alpha,T} - H_t \leq \frac{1}{n} \right\},$$
$$\eta_n = \inf \left\{ t \in [0,\infty) \colon M_t^{\alpha,T} - H_t \geq n \text{ or } |X_t| \geq n \right\}$$

and $\theta_n = \rho_n \wedge \eta_n \wedge \tau_n \wedge n$. For n sufficiently large, Itô's formula implies

$$\begin{aligned} &u\big(M_{\theta_n}^{\alpha,T} - H_{\theta_n}, X_{\theta_n}\big) - u(T,x) \\ &= \int_0^{\theta_n} \Big(u_T\big(M_s^{\alpha,T} - H_s, X_s\big)\alpha_s + \big(\nabla_x u\big(M_s^{\alpha,T} - H_s, X_s\big)\big)^\top \cdot \sigma(X_s) \Big) \cdot dW_s \\ &\quad + \int_0^{\theta_n} \left(-h(X_s)\, u_T + \mathcal{L}u + \frac{|\alpha_s|^2}{2} u_{TT} + (\nabla_x u_T)^\top \cdot \sigma(X_s) \cdot \alpha_s^\top \right) \big(M_s^{\alpha,T} - H_s, X_s\big) ds. \end{aligned}$$

Hence,

$$
\begin{aligned}
u(T,x) &= \mathbb{E}\big[u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\big] - \mathbb{E}\left[\int_0^{\theta_n} \Big(u_T\alpha_s + (\nabla_x u)^\top \cdot \sigma(X_s)\Big)\left(M^{\alpha,T}_s - H_s, X_s\right)\cdot dW_s\right] \\
&\quad + \mathbb{E}\left[\int_0^{\theta_n} \left(h\left(X_s\right)u_T - \mathcal{L}u - \frac{|\alpha_s|^2}{2}u_{TT} - (\nabla_x u_T)^\top \cdot \sigma(X_s)\cdot \alpha_s^\top\right)\big(M^{\alpha,T}_s - H_s, X_s\big)ds\right] \\
&\geq \mathbb{E}\big[u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\big] - \mathbb{E}\left[\int_0^{\theta_n} \Big(u_T\alpha_s + (\nabla_x u)^\top \cdot \sigma(X_s)\Big)\left(M^{\alpha,T}_s - H_s, X_s\right)\cdot dW_s\right] \\
&\quad + \mathbb{E}\left[\int_0^{\theta_n}\!\left(h\left(X_s\right)u_T - \mathcal{L}u - \sup_{a\in\mathbb{R}^d}\left\{\frac{|a|^2}{2}u_{TT} + (\nabla_x u_T)^\top\!\cdot \sigma(X_s)\cdot a\right\}\right)\!\big(M^{\alpha,T}_s\!- H_s, X_s\big)ds\right].
\end{aligned}
$$

Note that $I_s := \big(u_T\alpha_s + (\nabla_x u)^\top \cdot \sigma(X_s)\big)\big(M^{\alpha,T}_s - H_s, X_s\big)$ satisfies $\mathbb{E}\big[\int_0^{\theta_n} |I_s|^2 ds\big] < \infty$ and hence, the stochastic integral $\int_0^{\theta_n} I_s \cdot dW_s$ has expectation 0 by Proposition 1.23, Chapter IV in [56]. Since u is a concave supersolution of (II.3.3) with $\big\{(T,x) : u_{TT}(T,x) = 0\big\} \subseteq \big\{(T,x) : \nabla_x u_T(T,x) = 0 \in \mathbb{R}^n\big\}$, we have

$$
u(T,x) \geq \mathbb{E}\big[u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\big].
$$

For $n \to \infty$, θ_n converge to $\tau = \tau^{\alpha,T} := \inf\big\{t \in [0,\infty) \colon M^{\alpha,T}_t \leq H_t\big\}$, almost surely. The pathwise continuity of X_s and $M^{\alpha,T}_s - H_s$ imply that

$$
M^{\alpha,T}_{\theta_n} - H_{\theta_n} \xrightarrow[n\to\infty]{} M^{\alpha,T}_\tau - H_\tau = 0 \quad \text{and} \quad X_{\theta_n} \xrightarrow[n\to\infty]{} X_\tau, \quad \mathbb{P}\text{−a.s.}
$$

We deduce that

$$
u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big) \xrightarrow[n\to\infty]{} u\big(0, X^x_\tau\big) = f\left(X_\tau\right)
$$

from the continuity of u on $[0,\infty) \times \mathbb{R}^n$ and the initial condition. (II.4.1) implies that

$$
\Big|u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\Big| \leq C\big(1 + M^{\alpha,T}_{\theta_n} - H_{\theta_n} + |X_{\theta_n}|^p\big) \leq C\big(1 + M^{\alpha,T}_{\theta_n} + |X_{\theta_n}|^p\big).
$$

Observe that $\big\{M^{\alpha,T}_{\theta_n}\big\}_{n\in\mathbb{N}}$ is uniformly integrable, because $M^{\alpha,T}_{\theta_n} = \mathbb{E}\big[M^{\alpha,T}_\infty \,\big|\, \mathcal{F}_{\theta_n}\big]$ for $\alpha \in \mathcal{A}$. By assumption $\{|X_{\theta_n}|^p\}_{n\in\mathbb{N}}$ is also uniformly integrable and, hence, so is $\big\{u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\big\}_{n\in\mathbb{N}}$. Then it follows that

$$
\lim_{n\to\infty} \mathbb{E}\left[u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\right] = \mathbb{E}\left[f\left(X_\tau\right)\right]. \tag{II.4.5}
$$

Therefore, we have for all $(\alpha_s)_{s\in[0,\infty)} \in \mathcal{A}$

$$
u(T,x) \geq \mathbb{E}\left[f\left(X_\tau\right)\right]
$$

and hence, $u(T,x) \geq U(T,x)$.

2a) Now assume that for $(T,x) \in (0,\infty) \times \mathbb{R}^n$

$$
\alpha^*_s := -\left(\mathbb{1}_{\{u_{TT}<0\}}\frac{(\nabla_x u_T)^\top \cdot \sigma(X_s)}{u_{TT}}\right)(M^*_s - H_s, X_s) \in \mathcal{A},
$$

where

$$
dM^*_s = -\mathbb{1}_{\{M^*_s > H_s\}}\left(\mathbb{1}_{\{u_{TT}<0\}}\frac{(\nabla_x u_T)^\top \cdot \sigma(X_s)}{u_{TT}}\right)(M^*_s - H_s, X_s)\cdot dW_s, \quad M^*_0 = T.
$$

Let $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for α^*, define the stopping times ρ_n, η_n and θ_n as in the first part of the proof and apply Itô's formula to $u(M^*_{\theta_n} - H_{\theta_n}, X_{\theta_n})$. Notice that the definition of α^* implies that

$$u(T,x) = \mathbb{E}\big[u\big(M^*_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\big].$$

Finally, we have

$$\big|u\big(M^*_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\big| \le C\big(1 + M^*_{\theta_n} + |X_{\theta_n}|^p\big),$$

showing that $\big\{u\big(M^{\alpha,T}_{\theta_n} - H_{\theta_n}, X_{\theta_n}\big)\big\}_{n\in\mathbb{N}}$ is uniformly integrable. Taking the limit $n\to\infty$ results in

$$u(T,x) = \mathbb{E}\left[f\left(X_\tau\right)\right] \le U(T,x),$$

which implies the second claim.

2b) For the last part notice that u_T is non-increasing in T for fixed x by the concavity of $T\mapsto u(T,x)$. Thus, by Lemma II.1.3 we have that $m^* = \widetilde{T}(x)\wedge T$ and the corresponding optimal α^* given in *2a)* are optimal for $V(T,x)$. □

Remark II.4.2. Suppose that $h(y) \ge C\max\big\{|b(y)|, \|\sigma(y)\|^2\big\}$, $y\in\mathbb{R}$, for some $C\in(0,\infty)$, where $\|.\|$ denotes the Frobenius norm, i.e. $\|\sigma(x)\|^2 = \sum_{i=1}^n\sum_{j=1}^d \sigma^2_{ij}(x)$. Then the family $\mathcal{X}^1_\tau = \{|X^x_\vartheta| \colon \vartheta\ (\mathcal{F}_t)\text{-stopping time}, \vartheta\le\tau \text{ a.s.}\}$ is uniformly integrable for every $\tau\in\mathcal{S}(T)$, $T\in(0,\infty)$. Indeed, let $\tau\in\mathcal{S}(T)$ and ϑ be an $(\mathcal{F}_t)$-stopping time with $\vartheta\le\tau$ a.s., then

$$|X^x_\vartheta| \le \sup_{0\le t\le\tau}|X^x_t|.$$

We show that $\sup_{0\le t\le\tau}|X^x_t|$ is integrable and, therefore, $\mathcal{X}^1_\tau$ is uniformly integrable. Observe that

$$\begin{aligned}\sup_{0\le t\le\tau}|X^x_t| &\le |x| + \sup_{0\le t\le\tau}\left|\int_0^t \sigma(X^x_s)\cdot dW_s\right| + \sup_{0\le t\le\tau}\left|\int_0^t b(X^x_s)ds\right|\\ &\le |x| + \sup_{0\le t\le\tau}\left|\int_0^t \sigma(X^x_s)\cdot dW_s\right| + \int_0^\tau |b(X^x_s)|ds.\end{aligned}$$

Applying the Burkholder-Davis-Gundy inequality and the assumption that $\|\sigma(y)\|^2 \le \frac{1}{C}h(y)$ results in

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le t\le\tau}\left|\int_0^t\sigma(X^x_s)\cdot dW_s\right|\right] &\le C_1\mathbb{E}\left[\left(\int_0^\tau \|\sigma(X^x_s)\|^2ds\right)^{1/2}\right] \le C_1\left(1+\mathbb{E}\left[\int_0^\tau\|\sigma(X^x_s)\|^2ds\right]\right)\\ &\le C_1\left(1+\frac{1}{C}\mathbb{E}\left[\int_0^\tau h(X^x_s)ds\right]\right) = C_1\left(1+\frac{T}{C}\right),\end{aligned}$$

where C_1 is the constant arising in the Burkholder-Davis-Gundy inequality. Moreover, using $|b(y)|\le\frac{1}{C}h(y)$ we obtain

$$\mathbb{E}\left[\int_0^\tau |b(X^x_s)|ds\right] \le \frac{1}{C}\mathbb{E}\left[\int_0^\tau h(X^x_s)ds\right] \le \frac{T}{C}.$$

Hence, $\sup_{0\le t\le\tau}|X^x_t|$ is integrable and $\mathcal{X}^1_\tau$ is uniformly integrable for all $\tau\in\mathcal{S}(T)$.

Remark II.4.3. Let $X^x_t = x + W_t$, $t\in[0,\infty)$, and $h(y)=1, y\in\mathbb{R}^d$, and assume that the unique optimal control $\alpha^*\in\mathcal{A}$ for $U(T,x)$ is given by $\alpha^*_s = -2(X^x_s)^\top/d$ for every $(T,x)\in(0,\infty)\times\mathbb{R}^n$. Hence, the corresponding optimal stopping time is the first exit time of the ball around 0 with radius $R_x = \sqrt{dT+x^2}$ (cf. with Example II.2.7). In view of Equation (II.4.2) one can argue that $\nabla_x U_T(T,x)/U_{TT}(T,x) = 2x/d$ and the PDE (II.3.3) simplifies to

$$U_T(T,x) - \frac{1}{2}\Delta_x U(T,x) + \frac{1}{d}\,x^\top\cdot\nabla_x U_T(T,x) = 0.$$

Corollary II.4.4. *If the assumptions of Theorem II.4.1 are satisfied, then V is a solution to*

$$\min\left\{V_T(T,x),\, h(x)V_T(T,x) - \mathcal{L}V(T,x) + \frac{\left|\sigma^\top(x)\cdot\nabla_x V_T(T,x)\right|^2}{2V_{TT}(T,x)}\right\} = 0 \tag{II.4.6}$$

on $(0,\infty)\times\mathbb{R}^n \backslash \partial\big\{(T,x)\in(0,\infty)\times\mathbb{R}^n\colon V_T(T,x)=0\big\}$ *with initial condition* $V(0,x)=f(x)$, *where* ∂D *denotes the boundary of a set* $D\subseteq(0,\infty)\times\mathbb{R}^n$.

Proof. Recall that $V(T,x)=U\big(T\wedge\widetilde{T}(x),x\big)$ by Lemma II.1.3, because U is concave in T and $U(T,x)\in\mathbb{R}$ for all $T\in(0,\infty)$, $x\in\mathbb{R}^n$. In particular, V_T is continuous on $(0,\infty)\times\mathbb{R}^n$.

On $B_1 := \big\{(T,x)\in(0,\infty)\times\mathbb{R}^n\colon V_T(T,x)>0\big\}$ the functions U and V coincide by (II.4.4). Thus, $V\in\mathcal{C}^2(B_1)$ and V satisfies (II.4.6). On the open set $B_2 := \big\{(T,x)\colon V_T(T,x)=0\big\}\backslash \partial\big\{(T,x)\colon V_T(T,x)=0\big\}$ the value function V is constant in T for fixed $x\in\mathbb{R}^n$. More precisely, we have $V_T(T,x)=V_{TT}(T,x)=V_{Tx_i}(T,x)=0$, $1\le i\le n$. Moreover, we have

$$-\mathcal{L}V(T,x) = -\mathcal{L}U(T(x),x) = -\frac{\left|\sigma^\top(x)\cdot\nabla_x U_T(T(x),x)\right|^2}{2U_{TT}(T(x),x)} \geq 0.$$

Here we use that $U_T(T(x),x)=0$ and that U solves (II.3.3). Thus, V solves (II.4.6). □

Remark II.4.5. If the assumptions of Theorem II.4.1 are satisfied, then in general V is not a classical solution to (II.4.6) on $(0,\infty)\times\mathbb{R}^n$, see Example II.4.7 below.

We next apply Theorem II.4.1 in order to determine the optimal stopping time for various examples.

Example II.4.6 (Maximizing the Euclidean norm of a Brownian motion)**.** Let $f(y)=|y|$, $h(y)=1$, $y\in\mathbb{R}^d$, and $X_t^x = x+W_t$, $t\in[0,\infty)$, be a d-dimensional Brownian motion starting in $x\in\mathbb{R}^d$. Then the value function V of the optimal control problem $(\mathcal{V})$ is given by

$$V(T,x)=\sqrt{dT+|x|^2}.$$

An optimal control for $V(T,x)$ is $(\alpha_s^*,m^*)=(-2(X_s^x)^\top,T)$ and the corresponding optimal stopping time is the first exit time of the ball around 0 with radius $\sqrt{dT+|x|^2}$.

To see this, we first show with Theorem II.4.1 that $u(T,x):=\sqrt{dT+|x|^2}$ is the value function of the optimal control problem $(\mathcal{U})$ with optimal control $\alpha_s^* = -2(X_s^x)^\top$.

It is obvious that $u\in\mathcal{C}^2\big((0,\infty)\times\mathbb{R}^d\big)\cap\mathcal{C}\big([0,\infty)\times\mathbb{R}^d\big)$, $u(0,x)=|x|=f(x)$ and that $T\mapsto\sqrt{dT+|x|^2}$ is concave for fixed $x\in\mathbb{R}^d$. Moreover, $u_{TT}<0$ on $(0,\infty)\times\mathbb{R}^d$,

$$|u(T,x)| \le \sqrt{dT}+|x| \le d\big(1+T+|x|\big)$$

and u satisfies the PDE (II.3.3):

$$\begin{aligned}
&h(x)u_T(T,x) - \frac{1}{2}\Delta_x u(T,x) + \frac{\left|\nabla_x u_T(T,x)\right|^2}{2u_{TT}(T,x)}\\
&\quad= \frac{d}{2\big(dT+|x|^2\big)^{\frac12}} - \frac{1}{2}\sum_{i=1}^d\left(\frac{1}{\big(dT+|x|^2\big)^{\frac12}} - \frac{x_i^2}{\big(dT+|x|^2\big)^{\frac32}}\right) - \frac{1}{2d^2}\sum_{i=1}^d \frac{d^2x_i^2}{\big(dT+|x|^2\big)^{\frac32}} = 0.
\end{aligned}$$

Define

$$\alpha_s^* = -\left(\frac{\nabla_x u_T}{u_{TT}}\right)(M_s^*-s,X_s^x) = -\frac{2(X_s^x)^\top}{d}, \qquad M_t^* = T+\int_0^t \alpha_s^*\mathbb{1}_{\{M_s^*>s\}}\cdot dW_s.$$

Then, $(\alpha_s^*)_{s\in[0,\infty)}\in L^2_{loc}(W)$ and

$$M_t^* = T+\frac{|x|^2}{d} - \frac{|X_{t\wedge\rho}^x|^2}{d} + t\wedge\rho,$$

where $\rho = \rho(T,x) = \inf\{t \in [0,\infty) \colon M_t^* \le t\} = \inf\big\{t \in [0,\infty) \colon |X_t^x| \ge \sqrt{dT + |x|^2}\big\}$. In addition, Example II.2.7 implies that $\alpha^* \in \mathcal{A}$. Since $\mathcal{X}_\tau^1$ is uniformly integrable for all $\tau \in \mathcal{S}(T)$ by Remark II.4.2, the verification theorem shows that $u(T,x)$ is the value function of the optimal control problem ($\mathcal{U}$). Moreover, $u_T > 0$ implies that $V(T,x) = U(T,x)$ with optimal control (α^*, T). Using the one-to-one correspondence established in Proposition II.2.4, the optimal stopping time in (II.1.2) is given by $\rho(T,x)$.

Example II.4.7 (A value function which is strictly decreasing in time)**.** Here we present an example, where the value function U is a classical solution to (II.3.3), but V is not in $\mathcal{C}^2\big((0,\infty)\times\mathbb{R}\big)$.

Consider $f(y) = -y^4 + y^2$, $h(y) = 1$, $y \in \mathbb{R}$, and let $X_t^x = x + W_t$, $t \in [0,\infty)$, be a one-dimensional Brownian motion starting in $x \in \mathbb{R}$. We will show that

$$u(T,x) := -x^4 + x^2 + T - T^2 - 2x^2 T$$

is the value function of ($\mathcal{U}$) with optimal control $\alpha_s^* = -2X_s^x, s \in [0,\infty)$, whereas the value function V of the optimal control problem ($\mathcal{V}$) is given by

$$V(T,x) = \begin{cases} -x^4 + x^2 + T - T^2 - 2x^2 T, & \text{if } x^2 \le \frac{1}{2},\ T < \frac{1}{2} - x^2, \\ \frac{1}{4}, & \text{if } x^2 \le \frac{1}{2},\ T \ge \frac{1}{2} - x^2, \\ -x^4 + x^2, & \text{if } x^2 > \frac{1}{2}. \end{cases} \tag{II.4.7}$$

The optimal control is given by (α^*, T^*), where $T^* = \big(\big(\frac{1}{2} - x^2\big) \wedge T\big)^+$.

The function u is in $\mathcal{C}^2\big((0,\infty)\times\mathbb{R}\big) \cap \mathcal{C}\big([0,\infty)\times\mathbb{R}\big)$, $u(0,x) = -x^4 + x^2 = f(x)$, u is concave in T with $u_{TT} = -2 < 0$ on $(0,\infty)\times\mathbb{R}$ and u satisfies the PDE (II.3.3). Notice that u does not satisfy the growth condition (II.4.1), which guarantees that (II.4.5) holds. Nevertheless, we can perform a verification: We show that (II.4.5) is still valid in this example. We have

$$|u(T,x)| \le C\big(1 + T^2 + x^4\big)$$

for some $C \in (0,\infty)$. Lemma A.1.2 allows to restrict $\mathcal{T}(T)$, $T \in [0,\infty)$, to the set of all square integrable stopping times, i.e.

$$U(T,x) = \sup_{\tau \in \mathcal{T}(T) \text{ with } \mathbb{E}[\tau^2] < \infty} \mathbb{E}[f(X_\tau^x)].$$

Let $\tau \in \mathcal{T}(T)$ be square integrable with corresponding control α. Then $\sup_{0\le s\le\tau} M_s^{\alpha,T}$ is square integrable by Doob's L^2 inequality. Moreover, it holds that $\mathbb{E}[\sup_{0\le s\le\tau} |X_\tau^x|^4] \le C\big(|x|^2 + \mathbb{E}[\tau^2]\big) < \infty$, $C \in (0,\infty)$, see the proof of Lemma A.1.2. Thus, similar arguments as in the proof of Theorem II.4.1 imply that (II.4.5) holds and hence we have $u(T,x) \ge U(T,x)$. The details are given in the proof of Lemma A.1.2. Furthermore, we have $u(T,x) = \mathbb{E}[f(X_{\tau^*}^x)]$, where τ^* denotes the first exit time of the interval $(-\sqrt{T + x^2}, \sqrt{T + x^2})$. Observe that $\mathbb{E}[\tau^*] = T$ by Example II.2.7. Therefore, u is the value function of the optimal control problem ($\mathcal{U}$). By Theorem II.4.1 the value function V is given by (II.4.7) and thus, V is not in $\mathcal{C}^2\big((0,\infty)\times\mathbb{R}\big)$.

Remark II.4.8. We remark that in Example II.4.6 and II.4.7 one can perform a verification without using Theorem II.4.1: Indeed, observe that for a d-dimensional Brownian motion $(X_t^x)_{t\in[0,\infty)}$ we have for $\tau \in \mathcal{T}(T)$ and $t \in [0,\infty)$ that

$$\mathbb{E}\big[|X_{\tau\wedge t}^x|^2\big] = |x|^2 + d\mathbb{E}[\tau \wedge t] \le |x|^2 + d\mathbb{E}[\tau] = |x|^2 + dT.$$

Thus, $(X_{\tau\wedge t}^x)_{t\in[0,\infty)}$ is bounded in $L^2(\Omega)$. Hence, $\lim_{t\to\infty} \mathbb{E}\big[|X_{\tau\wedge t}^x|^2\big] = \mathbb{E}\big[|X_\tau^x|^2\big]$ by the L^2 martingale convergence theorem. Monotone convergence implies $\lim_{t\to\infty} \mathbb{E}[\tau \wedge t] = \mathbb{E}[\tau]$. Therefore,

$$\mathbb{E}\big[|X_\tau^x|^2\big] = |x|^2 + dT. \tag{II.4.8}$$

Assume that $f(y) = g(a|y-c|^2)$, where $a \in \mathbb{R}$, $c \in \mathbb{R}^d$, $g\colon D \subseteq \mathbb{R} \to \mathbb{R}$, $z \mapsto g(z)$ is concave. Then Jensen's inequality and (II.4.8) imply that

$$\mathbb{E}[f(X_\tau^x)] \le g\big(a\mathbb{E}\big[|X_\tau^x - c|^2\big]\big) = g\left(a\big(|x-c|^2 + dT\big)\right)$$

for all $\tau \in \mathcal{T}(T)$ and thus, $U(T,x) \le g\left(a\big(|x-c|^2+dT\big)\right)$. The stopping time

$$\rho(T,x) := \inf\left\{t \in [0,\infty)\colon |X_t^x - c| \ge \sqrt{|x-c|^2 + dT}\right\}$$

has expectation T (cf. Chapter 4.2.E in [37]) and hence,

$$U(T,x) \ge \mathbb{E}\left[f\big(X_{\rho(T,x)}^x\big)\right] = \mathbb{E}\left[g\big(a|X_{\rho(T,x)}^x - c|^2\big)\right] = g\left(a\big(|x-c|^2+dT\big)\right).$$

Therefore,

$$U(T,x) = g\left(a\big(|x-c|^2+dT\big)\right).$$

In Example II.4.6 and II.4.7 we have $f(y) = |y| = g_1(|y|^2)$ and $f(y) = -y^4 + y^2 = g_2(y^2)$, respectively, where $g_1(z) = \sqrt{z}$, $z \ge 0$, and $g_2(z) = -z^2 + z$, $z \ge 0$. Hence, the value functions are given by

$$U(T,x) = g_1\big(|x|^2 + dT\big) = \sqrt{|x|^2 + dT}$$

and

$$U(T,x) = g_2\big(x^2 + T\big) = -x^4 - T^2 - 2x^2T + x^2 + T,$$

respectively.

Example II.4.9 (The optimal stopping time is not a first hitting time of two points)**.** In the previous two examples the optimal stopping times are first exit times of a ball and an interval, respectively. We now present an example where this is not the case.

For $f(y) = y^2\mathbb{1}_{\{|y|\ge 1\}}$, $h(y) = 1, y \in \mathbb{R}$, and a one-dimensional Brownian motion $X_t^x = x + W_t$, $t \in [0,\infty)$, which starts in $x \in \mathbb{R}$, the value functions U and V of the optimal control problems $(\mathcal{U})$ and $(\mathcal{V})$ both coincide with the function u which is given by

$$u(T,x) = \begin{cases} \dfrac{T}{T + (1-|x|)^2}, & T < |x|(1-|x|), \\ T + x^2, & T \ge |x|(1-|x|). \end{cases}$$

Note that u is only in $\mathcal{C}^{1,1}\big((0,\infty)\times\mathbb{R}\big)$, because U_T and U_x are neither differentiable with respect to T nor with respect to x in $\big(|x|(1-|x|), x\big)$ for $0 < |x| < 1$. Therefore, we slightly modify the arguments used in the proof of Theorem II.4.1 to show that u is indeed the value function of the optimal control problem $(\mathcal{U})$.

Let $\mathcal{R} = \{(t,x) \in (0,\infty)\times\mathbb{R}\colon t < |x|(1-|x|)\}$. In order to verify that $U = u$ on $\mathcal{R}$, let $(T,x) \in \mathcal{R}$, $\alpha \in \mathcal{A}$ and define the stopping times

$$\begin{aligned} \rho &= \inf\left\{t \in [0,\infty)\colon \big(M_t^{\alpha,T} - t, X_t^x\big) \notin \mathcal{R}\right\}, \\ \theta_n &= \inf\left\{t \in [0,\infty)\colon M_t^{\alpha,T} - t \le \frac{T}{2n}\right\} \wedge \tau_n \wedge n, \end{aligned}$$

where $(\tau_n)_{n\in\mathbb{N}}$ is a localizing sequence for α. Observe that $u \in \mathcal{C}^2(\mathcal{R})$. Moreover, the definition of θ_n and ρ imply that the stochastic integral $N_t = \int_0^t (u_x + \alpha_s u_T)\big(M_s^{\alpha,T} - s, X_s^x\big)dW_s$ is a martingale on $[0, \theta_n \wedge \rho]$ by Proposition 1.23, Chapter IV in [56]. Hence, by Itô's formula

$$\begin{aligned} u(T,x) &= \mathbb{E}\big[u\big(M^{\alpha,T}_{\theta_n\wedge\rho} - \theta_n \wedge \rho, X^x_{\theta_n\wedge\rho}\big)\big] \\ &\quad + \mathbb{E}\left[\int_0^{\theta_n\wedge\rho} \left(u_T - \mathcal{L}u - \frac{|\alpha_s|^2}{2} u_{TT} - \alpha_s u_{Tx}\right)(M^{\alpha,T}_s - s, X^x_s) ds\right] \\ &\geq \mathbb{E}\left[u\big(M^{\alpha,T}_{\theta_n\wedge\rho} - \theta_n \wedge \rho, X^x_{\theta_n\wedge\rho}\big) + \int_0^{\theta_n\wedge\rho} \left(u_T - \mathcal{L}u + \frac{u_{Tx}^2}{2u_{TT}}\right)(M^{\alpha,T}_s - s, X^x_s) ds\right] \\ &= \mathbb{E}\big[u\big(M^{\alpha,T}_{\theta_n\wedge\rho} - \theta_n \wedge \rho, X^x_{\theta_n\wedge\rho}\big)\big]. \end{aligned}$$

In the last step we use that u is a solution to (II.3.3) on $\mathcal{R}$. The stopping times θ_n converge to $\tau = \tau^{\alpha,T} := \inf\big\{t \in [0,\infty) \colon M^{\alpha,T}_t \leq t\big\}$ as $n \to \infty$. Recall that $M^{\alpha,T}_\tau = \tau$. Hence, by dominated convergence

$$\begin{aligned} u(T,x) &\geq \mathbb{E}\big[u\big(M^{\alpha,T}_{\tau\wedge\rho} - \tau\wedge\rho, X^x_{\tau\wedge\rho}\big)\big] \\ &= \mathbb{E}\big[\mathbb{1}_{\{\tau\leq\rho\}} u\big(M^{\alpha,T}_\tau - \tau, X^x_\tau\big) + \mathbb{1}_{\{\rho<\tau\}} u\big(M^{\alpha,T}_\rho - \rho, X^x_\rho\big)\big] \\ &= \mathbb{E}\big[\mathbb{1}_{\{\tau\leq\rho\}} f(X^x_\tau) + \mathbb{1}_{\{\rho<\tau\}} \left(M^{\alpha,T}_\rho - \rho + (X^x_\rho)^2\right)\big]. \end{aligned}$$

Recall that for $\alpha \in \mathcal{A}$ the process $\big(M^{\alpha,T}_{t\wedge\tau}\big)_{t\in[0,\infty]}$ is a martingale. Moreover, $\big((X^x_{t\wedge\tau})^2 - t\wedge\tau\big)_{t\in[0,\infty]}$ is a martingale, because its quadratic variation $(\tau \wedge t)_{t\in[0,\infty]}$ is integrable (see Proposition 1.23, Chapter IV in [56]). Therefore,

$$\begin{aligned} u(T,x) &\geq \mathbb{E}\left[\mathbb{1}_{\{\tau\leq\rho\}} f(X^x_\tau) + \mathbb{1}_{\{\rho<\tau\}} \mathbb{E}\left[M^{\alpha,T}_\tau - \tau + (X^x_\tau)^2 \,\middle|\, \mathcal{F}_\rho\right]\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{\tau\leq\rho\}} f(X^x_\tau) + \mathbb{1}_{\{\rho<\tau\}} (X^x_\tau)^2\right] \\ &\geq \mathbb{E}\left[\mathbb{1}_{\{\tau\leq\rho\}} f(X^x_\tau) + \mathbb{1}_{\{\rho<\tau\}} f(X^x_\tau)\right] \\ &= \mathbb{E}[f(X^x_\tau)]. \end{aligned}$$

To summarize, it holds that $u(T,x) \geq U(T,x)$ on $\mathcal{R}$. For $(T,x) \in \mathcal{R}$ the stopping time

$$\begin{aligned} \rho(T,x) &= \inf\left\{t \in [0,\infty) \colon X^x_t \notin \left(x - \frac{T}{1-x}, 1\right)\right\}, \qquad \text{if } x > 0, \\ \rho(T,x) &= \inf\left\{t \in [0,\infty) \colon X^x_t \notin \left(-1, x + \frac{T}{1+x}\right)\right\}, \quad \text{if } x < 0 \end{aligned}$$

satisfies $\mathbb{E}[\rho(T,x)] = T$ by Example II.2.8. Hence,

$$u(T,x) = \mathbb{E}\big[f\big(X^x_{\rho(T,x)}\big)\big] \leq U(T,x).$$

Hence, $u = U$ on $\mathcal{R}$ and $\rho(T,x)$ is optimal for $(T,x) \in \mathcal{R}$. For $(T,x) \in \mathcal{R}^c$, note that $f(y) \leq y^2$ and, thus, by (II.4.8)

$$U(T,x) \leq \sup_{\tau\in\mathcal{T}(T)} \mathbb{E}\left[(X^x_\tau)^2\right] = x^2 + T = u(T,x).$$

For the reverse inequality let $\vartheta = \inf\{t \in [0,\infty) \colon (T-t, X^x_t) \in \mathcal{R}\}$ and define

$$\alpha^*_s = \big(-2X^x_s + \operatorname{sgn}(X^x_\vartheta)\big)\mathbb{1}_{\{s\geq\vartheta\}},$$

where $\operatorname{sgn}(a) = 1$ if $a \in (0,\infty)$, $\operatorname{sgn}(a) = -1$ if $a \in (-\infty, 0)$ and $\operatorname{sgn}(0) = 0$. Thus,

$$M^*_t = \begin{cases} T, & \text{on } \{T \leq \vartheta\} \cup \{t \leq \vartheta < T\}, \\ -(X^x_{t\wedge\tau})^2 + (t\wedge\tau) + \operatorname{sgn}(X^x_\vartheta) X^x_{t\wedge\tau}, & \text{on } \{\vartheta < T\} \cap \{t > \vartheta\}, \end{cases}$$

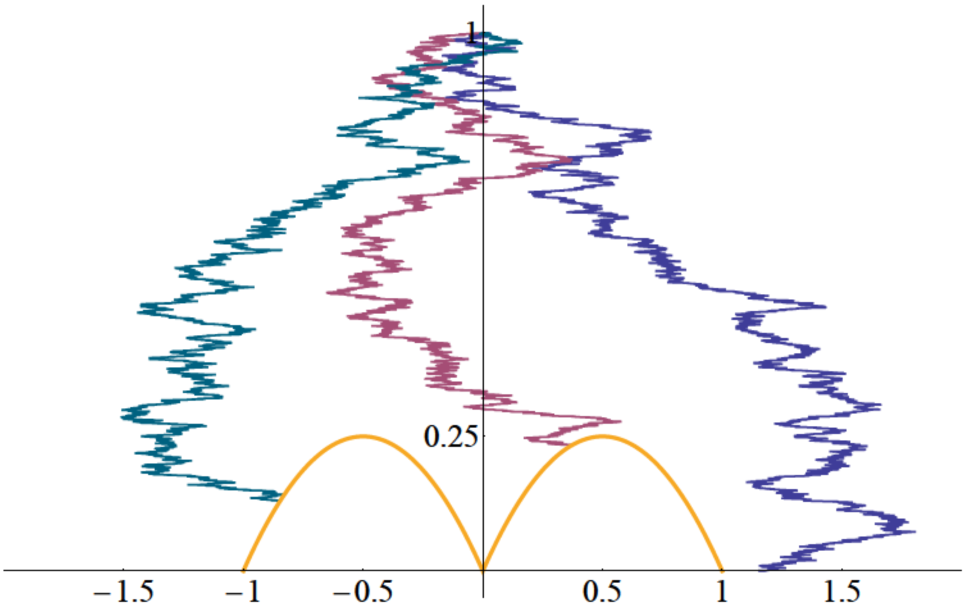


Figure II.1: The figure depicts three realizations of the pair $(X_t^0, M_t^{*,1} - t)$ in Example II.4.9. The y-axis is the conditional expected remaining time.

where

$$\tau = \tau_\vartheta = \mathbb{1}_{\{\vartheta \geq T\}} T + \mathbb{1}_{\{\vartheta < T\}} \left(\vartheta + \tau^x_{0,\,\mathrm{sgn}(X^x_\vartheta)}(\vartheta)\right)$$

with $\tau^x_{a,b}(\vartheta) = \inf\{t \in [0,\infty) \colon X^x_{\vartheta+t} \notin (a \wedge b, a \vee b)\}$, $a, b \in \mathbb{R}$. Observe that $\tau_\vartheta = \inf\{t \in [0,\infty) \colon M^*_t \leq t\}$. The strong Markov property implies

$$\begin{aligned}
\mathbb{E}[\tau_\vartheta] &= \mathbb{E}\left[\mathbb{1}_{\{\vartheta \geq T\}} T + \mathbb{1}_{\{\vartheta < T\}} \left(\vartheta + \tau^x_{0,\,\mathrm{sgn}(X^x_\vartheta)}(\vartheta)\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{\vartheta \geq T\}} T + \mathbb{1}_{\{\vartheta < T\}} \left(\vartheta + \mathbb{E}\left[\tau^x_{0,\,\mathrm{sgn}(X^x_\vartheta)}(\vartheta) \,\middle|\, \mathcal{F}_\vartheta\right]\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{\vartheta \geq T\}} T + \mathbb{1}_{\{\vartheta < T\}} \left(\vartheta + (1 - |X^x_\vartheta|)\,|X^x_\vartheta|\,\right)\right] \\
&= T,
\end{aligned}$$

where we use in the last step that $T - \vartheta = |X^x_\vartheta|\,(1 - |X^x_\vartheta|)$ by the definition of ϑ. Therefore, $\mathbb{E}[\tau_\vartheta] = M^*_0$ and thus $\alpha^* \in \mathcal{A}$. Combining the arguments given in the proof of the second part of Theorem II.4.1 with first stopping at ϑ and using that $\left(M^*_{t\wedge\tau_\vartheta}\right)_{t\in[0,\infty]}$ and $\left((X^x_{t\wedge\tau_\vartheta})^2 - t\wedge\tau_\vartheta\right)_{t\in[0,\infty]}$ are martingales, results in

$$\begin{aligned}
u(T,x) &= \mathbb{E}\left[M^*_{\tau_\vartheta\wedge\vartheta} - (\tau_\vartheta \wedge \vartheta) + (X^x_{\tau_\vartheta\wedge\vartheta})^2\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{\tau_\vartheta \leq \vartheta\}} f(X^x_{\tau_\vartheta}) + \mathbb{1}_{\{\vartheta < \tau_\vartheta\}} \left(M^*_\vartheta - \vartheta + (X^x_\vartheta)^2\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{\tau_\vartheta \leq \vartheta\}} f(X^x_{\tau_\vartheta}) + \mathbb{1}_{\{\vartheta < \tau_\vartheta\}} \left(X^x_{\tau_\vartheta}\right)^2\right],
\end{aligned}$$

where we use that on $\{\tau_\vartheta \leq \vartheta\} = \{\vartheta \geq T\}$ it holds that $\left|X^x_{\tau_\vartheta}\right| \in \{0\} \cup [1,\infty)$ and thus, $\left(X^x_{\tau_\vartheta}\right)^2 = f(X^x_{\tau_\vartheta})$. On $\{\vartheta < \tau_\vartheta\} = \{\vartheta < T\}$ we have by the definition of τ_ϑ that $X^x_{\tau_\vartheta} \in \{-1, 0, 1\}$, $\mathbb{P}$-a.s., which implies that $\left(X^x_{\tau_\vartheta}\right)^2 = f\left(X^x_{\tau_\vartheta}\right)$, and hence

$$u(T,x) = u(T,x) = \mathbb{E}\left[f\left(X^x_{\tau_\vartheta}\right)\right] \leq U(T,x).$$

Therefore, α^* is optimal in $(\mathcal{U})$ for $T \geq |x|(1-|x|)$ and the optimal stopping time in (II.1.3) is given by τ_ϑ. Notice that, for $T > |x|(1-|x|)$, the optimal strategy can be described by the words: "Do nothing until $\vartheta \wedge T$; after ϑ, provided $\vartheta < T$, control the process M in such a way that the space-time process $(X^x_s, M_s - s)$ stays on the graph of $x(1-x)$ if $X^x_\vartheta > 0$ resp. $-x(1+x)$ if $X^x_\vartheta < 0$." Figure II.1 illustrates possible paths the pair $(X^0_t, M^{*,1}_t - t)$ can take.

II.5 Two Families of Optimal Stopping Problems

In this section we derive optimal stopping problems with non-constant constraint functions h and their value functions for the one-dimensional case $d = n = 1$. Here we allow for arbitrary drift, diffusion and constraint functions within our assumptions. We impose that a classical solution u of the PDE (II.3.3) has a specific form. Then the payoff function f for an optimal stopping problem with value function u is given by $f(x) = u(0, x)$. Finally, we apply Theorem II.4.1 to verify that u is indeed the value function of the optimal stopping problem (II.1.3).

II.5.1 Additive Structure

First we use the ansatz $u(T, x) = g\big(\ell(T) + k(x)\big)$ for $(T, x) \in [0, \infty) \times \mathbb{R}$, where $\ell \in \mathcal{C}^2\big((0, \infty)\big) \cap \mathcal{C}\big([0, \infty)\big)$, $k \in \mathcal{C}^2(\mathbb{R})$ and $g \in \mathcal{C}^2\big(\mathring{A}\big) \cap \mathcal{C}(A)$, where $A = \{\ell(T) + k(x) \colon (T, x) \in [0, \infty) \times \mathbb{R}\}$ and $\mathring{A}$ denotes the interior of A. In addition, we assume that

$$u_{TT}(T, x) = \big(\ell'(T)\big)^2 g''\big(\ell(T) + k(x)\big) + \ell''(T) g'\big(\ell(T) + k(x)\big) < 0 \tag{II.5.1}$$

for all $(T, x) \in (0, \infty) \times \mathbb{R}$. Then the PDE (II.3.3) is given by

$$\begin{aligned} h(x)\ell'(T)g'\big(\ell(T) + k(x)\big) - \frac{\sigma^2(x)}{2}\Big[k''(x)g'\big(\ell(T) + k(x)\big) + \big(k'(x)\big)^2 g''\big(\ell(T) + k(x)\big)\Big] \\ - b(x)k'(x)g'\big(\ell(T) + k(x)\big) + \frac{\sigma^2(x)\big[k'(x)\ell'(T)g''\big(\ell(T) + k(x)\big)\big]^2}{2\big[\big(\ell'(T)\big)^2 g''\big(\ell(T) + k(x)\big) + \ell''(T)g'\big(\ell(T) + k(x)\big)\big]} = 0. \end{aligned} \tag{II.5.2}$$

To improve readability we omit the argument $\ell(T) + k(x)$ of the functions g, g' and g'' in the following. Using the assumption (II.5.1) we rewrite (II.5.2) as

$$\begin{aligned} g'g''\left[\big(\ell'(T)\big)^2\left\{h(x)\ell'(T) - \frac{\sigma^2(x)}{2}k''(x) - b(x)k'(x)\right\} - \frac{\sigma^2(x)}{2}\ell''(T)\big(k'(x)\big)^2\right] \\ + \ell''(T)\big(g'\big)^2\left[h(x)\ell'(T) - \frac{\sigma^2(x)}{2}k''(x) - b(x)k'(x)\right] = 0. \end{aligned} \tag{II.5.3}$$

We now focus on functions ℓ and k such that both summands in (II.5.3) equal 0 for all $(T, x) \in (0, \infty) \times \mathbb{R}$. We assume that $\ell''(T) = 0$ and

$$h(x)\ell'(T) - \frac{\sigma^2(x)}{2}k''(x) - b(x)k'(x) = 0$$

for all $T \in (0, \infty)$ and $x \in \mathbb{R}$. Thus, $\ell(T) = aT + \hat{a}$, $a, \hat{a} \in \mathbb{R}$. In the following we require that $a \neq 0$, otherwise (II.5.1) does not hold. The function k then satisfies

$$a\,h(x) - \frac{\sigma^2(x)}{2}k''(x) - b(x)k'(x) = 0, \quad x \in \mathbb{R}. \tag{II.5.4}$$

Hence,

$$k(x) = C_1 + \int_0^x \left\{\exp\left(-\int_0^z \frac{2b(w)}{\sigma^2(w)}dw\right)\left[\int_0^z \frac{2ah(w)}{\sigma^2(w)}\exp\left(\int_0^w \frac{2b(y)}{\sigma^2(y)}dy\right)dw + C_2\right]\right\}dz,$$

where $C_1, C_2 \in \mathbb{R}$. Without loss of generality we assume that $\hat{a} = 0$. Otherwise we replace C_1 by $C_1 + \hat{a}$. Observe that $u_{TT}(T, x) = a^2 g''\big(\ell(T) + k(x)\big)$. Therefore, if g is strictly concave on $A = \{aT + k(x) \colon T \in [0, \infty),\ x \in \mathbb{R}\}$, then u is strictly concave in T. Moreover, the payoff function f is given by $f(x) = u(0, x) = g\big(k(x)\big)$. Since we want to apply Theorem II.4.1 to verify that u is the value function of an optimal stopping problem, we choose g in such a way that the growth condition (II.4.1) is satisfied for $p = 1$. For different b, σ and h Table II.1 summarizes functions k

satisfying (II.5.4) and strictly concave functions g such that $g\big(aT + k(x)\big)$ fulfills (II.4.1) for $p = 1$. In addition, in all the examples the constraint function h satisfies Assumption (**A**), because it is bounded away from 0 and continuous. Moreover, it holds that $h(x) \geq C \max\{|b(x)|, \sigma^2(x)\}$ for all $x \in \mathbb{R}$, where $C \in (0, \infty)$. Therefore, in all examples of Table II.1 the family $\mathcal{X}^1_\tau$ is uniformly integrable for every $\tau \in \mathcal{S}(T)$, $T \in (0, \infty)$, by Remark II.4.2. Let

$$\alpha^*_s = -\sigma(X^x_s)\left(\frac{u_{Tx}}{u_{TT}}\right)(M^*_s - H^x_s, X^x_s) = -\frac{1}{a}\sigma(X^x_s)k'(X^x_s),$$
$$M^*_t = T + \int_0^t \alpha^*_s \mathbb{1}_{\{M^*_s > H^x_s\}} dW_s.$$

Then, $(\alpha^*_s)_{s\in[0,\infty)} \in L^2_{loc}(W)$. Let $\rho = \rho(T, x) = \inf\big\{t \in [0, \infty)\colon \frac{1}{a}k(X^x_t) \geq T + \frac{1}{a}k(x)\big\}$. Recall that k solves (II.5.4). Applying Itô's formula to $\frac{1}{a}k(X^x_{t\wedge\rho})$ we conclude that

$$\begin{aligned} M^*_t &= T + \frac{1}{a}k(x) - \frac{1}{a}k(X^x_{t\wedge\rho}) + \int_0^{t\wedge\rho} \frac{1}{a}\left(bk' + \frac{1}{2}\sigma^2 k''\right)(X^x_s)ds \\ &= T + \frac{1}{a}k(x) - \frac{1}{a}k(X^x_{t\wedge\rho}) + H^x_{t\wedge\rho} \end{aligned}$$

and $\rho = \inf\{t \in [0, \infty)\colon M^*_t \leq H^x_t\}$. Let $\theta_n = \inf\{t \in [0, \infty)\colon |X^x_t| \geq n\}$, $n \in \mathbb{N}$. Then monotone convergence and Itô's formula yield

$$\begin{aligned} \mathbb{E}\left[\int_0^{\rho(T,x)} h(X^x_s)ds\right] &= \lim_{n\to\infty} \mathbb{E}\left[\int_0^{\rho(T,x)\wedge\theta_n\wedge n} h(X^x_s)ds\right] \\ &= \frac{1}{a}\lim_{n\to\infty} \mathbb{E}\left[k\left(X^x_{\rho(T,x)\wedge\theta_n\wedge n}\right) - k(x) - \int_0^{\rho(T,x)\wedge\theta_n\wedge n} \sigma(X^x_s)k'(X^x_s)\,dW_s\right]. \end{aligned}$$

Observe that the stochastic integral is a martingale, because the integrand is bounded on $[0, \rho(T, x)\wedge \theta_n \wedge n]$. Moreover, in all examples of Table II.1 we have that $\big|k\big(X^x_{\rho(T,x)\wedge\theta_n\wedge n}\big)\big| \leq |a|T + |k(x)|$, $n \in \mathbb{N}$, by the definition of $\rho(T, x)$ and because k is bounded from below by 1 if $a > 0$ and from above by -1 if $a < 0$. Hence, dominated convergence and the definition of $\rho(T, x)$ imply

$$\mathbb{E}\left[H^x_{\rho(T,x)}\right] = \mathbb{E}\left[\int_0^{\rho(T,x)} h(X^x_s)ds\right] = \frac{1}{a}\mathbb{E}\left[k\left(X^x_{\rho(T,x)}\right) - k(x)\right] = T. \tag{II.5.5}$$

Thus, $\alpha^* \in \mathcal{A}$ for all examples of Table II.1 and, in particular, Assumption (**A**) implies that $\rho(T, x) < \infty$, $\mathbb{P}$-a.s. Therefore, in all examples the function $u(T, x) = g\big(aT + k(x)\big)$ is the value function of the optimal control problem $(\mathcal{U})$ with payoff function $g\big(k(x)\big)$ and constraint function h and with optimal control α^* by Theorem II.4.1. In the primal problem (II.1.3) an optimal stopping time for $u(T, x)$ is given by $\rho(T, x)$. In addition, u is strictly increasing in T. Thus, (α^*, T) and $\rho(T, x)$ are optimal in $(\mathcal{V})$ and (II.1.2), respectively, by Theorem II.4.1, 2b).

Remark II.5.1. Observe that in the examples on page 34 similar arguments as in Remark II.4.8 can be used to verify that the value function of the optimal stopping problem (II.1.3) is given by u without using Theorem II.4.1. Let $\theta_n = \inf\{t \in [0, \infty)\colon |X^x_t| \geq n\}$, $n \in \mathbb{N}$. Since $g\big(aT + k(x)\big)$ satisfies the growth condition (II.4.1) for $p = 1$ and $\mathcal{X}^1_\tau$ is uniformly integrable for every $\tau \in \mathcal{S}(T)$, we conclude that also $\big(g\big(k(X^x_{\tau\wedge\theta_n\wedge n})\big)\big)_{n\in\mathbb{N}}$ is uniformly integrable for every $\tau \in \mathcal{T}(T)$. Moreover, the concavity of g implies that

$$\begin{aligned} \mathbb{E}\left[g\big(k(X^x_\tau)\big)\right] &= \lim_{n\to\infty} \mathbb{E}\left[g\left(k\left(X^x_{\tau\wedge\theta_n\wedge n}\right)\right)\right] \leq \lim_{n\to\infty} g\left(\mathbb{E}\left[k\left(X^x_{\tau\wedge\theta_n\wedge n}\right)\right]\right) \\ &= \lim_{n\to\infty} g\left(\mathbb{E}\left[k(x) + \int_0^{\tau\wedge\theta_n\wedge n} (\sigma k')(X^x_s)\,dW_s + \int_0^{\tau\wedge\theta_n\wedge n} \Big(bk' + \frac{1}{2}\sigma^2 k''\Big)(X^x_s)ds\right]\right) \\ &= \lim_{n\to\infty} g\left(k(x) + \mathbb{E}\left[\int_0^{\tau\wedge\theta_n\wedge n} ah(X^x_s)ds\right]\right), \end{aligned}$$

where we use that the stochastic integral is a true martingale and that k is a solution to (II.5.4). The continuity of g and monotone convergence imply that

$$\mathbb{E}\left[g\big(k(X_\tau^x)\big)\right] \leq g\left(k(x) + a\mathbb{E}\left[\int_0^\tau h(X_s^x)ds\right]\right) = g\big(k(x) + a\mathbb{E}[H_\tau]\big) = g\big(k(x) + aT\big).$$

Hence, $U(T,x) \leq g\left(k(x) + aT\right)$. For the reverse inequality let

$$\rho(T,x) = \inf\left\{t \in [0,\infty) \colon \frac{1}{a}k(X_t^x) \geq \frac{1}{a}k(x) + T\right\}.$$

Then (II.5.5) implies that $\rho(T,x) \in \mathcal{T}(T)$. Therefore,

$$U(T,x) \geq \mathbb{E}\left[g\big(k\big(X_{\rho(T,x)}^x\big)\big)\right] = g\big(k(x) + aT\big).$$

To sum up, $U(T,x) = g\big(k(x) + aT\big)$ and an optimal stopping time is given by $\rho(T,x)$.

II.5.2 Multiplicative Structure

Now we assume that the variables T and x in a solution of (II.3.3) are separated. More precisely, we impose that a solution u of (II.3.3) is given by $u(T,x) = \ell(T)g(x)$, where $\ell \in \mathcal{C}^2\big((0,\infty)\big) \cap \mathcal{C}\big([0,\infty)\big)$ and $g \in \mathcal{C}^2(\mathbb{R})$. Then u solves

$$h(x)\ell'(T)g(x) - \frac{\sigma^2(x)}{2}\ell(T)g''(x) - b(x)\ell(T)g'(x) + \frac{\sigma^2(x)\big(\ell'(T)g'(x)\big)^2}{2\ell''(T)g(x)} = 0. \qquad \text{(II.5.6)}$$

Furthermore, we want to choose ℓ and g such that $u_{TT}(T,x) = \ell''(T)g(x) < 0$ for all $(T,x) \in (0,\infty) \times \mathbb{R}$. Then (II.5.6) can be rewritten as

$$2h(x)\ell'(T)\ell''(T)g^2(x) - \ell(T)\ell''(T)g(x)\left(\sigma^2(x)g''(x) + 2b(x)g'(x)\right) + \sigma^2(x)\big(\ell'(T)g'(x)\big)^2 = 0.$$

To obtain a differential equation for g independent of T, fix $c \in \mathbb{R}\backslash\{0\}$ and let $\ell(T) = \gamma e^{cT}$, $\gamma \in \mathbb{R}\backslash\{0\}$. This is the only non-constant function satisfying $\ell'\ell'' = c\ell\ell'' = c(\ell')^2$. Then g solves

$$2ch(x)g^2(x) + \sigma^2(x)\left(\big(g'(x)\big)^2 - g(x)g''(x)\right) - 2b(x)g(x)g'(x) = 0. \qquad \text{(II.5.7)}$$

A solution to (II.5.7) is given by

$$g(x) = C_1 \exp\left(\int_1^x \left[C_2 + \int_1^z \frac{2ch(y)}{\sigma^2(y)} \exp\left(\int_1^y \frac{2b(w)}{\sigma^2(w)}dw\right)dy\right] \exp\left(-\int_1^z \frac{2b(w)}{\sigma^2(w)}dw\right)dz\right),$$

where $C_1, C_2 \in \mathbb{R}$. Let $C_1 = 1$. For the payoff function f we have $f(x) = u(0,x) = \gamma g(x)$. Since $u_{TT}(T,x) = c^2e^{cT}f(x)$ it follows that u is strictly concave in T if and only if $f(x) < 0$ for all $x \in \mathbb{R}$. We set $\gamma = -1$ and thus require that g is a positive function. Then it follows that $f = -g$. Table II.2 summarizes functions g for different b, σ and h. In these examples u satisfies the growth condition (II.4.1) if and only if $c < 0$. Therefore, let $c < 0$. Furthermore, the constraint functions h are chosen such that Assumption (**A**) holds and that Remark II.4.2 can be applied. Let

$$\alpha_s^* = -\sigma(X_s^x)\left(\frac{u_{Tx}}{u_{TT}}\right)(M_s^* - H_s^x, X_s^x) = -\sigma(X_s^x)\frac{g'(X_s^x)}{cg(X_s^x)},$$

$$M_t^* = T + \int_0^t \alpha_s^* \mathbb{1}_{\{M_s^* > H_s^x\}} dW_s.$$

Then, $(\alpha_s^*)_{s\in[0,\infty)} \in L^2_{loc}(W)$. Itô's formula and (II.5.7) imply that

$$M_t^* = T + \frac{1}{c}\log\big(g(x)\big) - \frac{1}{c}\log\big(g(X_{t\wedge\rho}^x)\big) + H_{t\wedge\rho}^x,$$

where $\rho = \rho(T,x) = \inf\big\{t \in [0,\infty)\colon \log\big(g(X_t^x)\big) \le cT + \log\big(g(x)\big)\big\}$. Let $\theta_n = \inf\{t \in [0,\infty)\colon |X_t^x| \ge n\}$, $n \in \mathbb{N}$. Then monotone convergence, Itô's formula and the boundedness of X^x on $[0, \rho(T,x) \wedge \theta_n \wedge n]$ yield

$$\mathbb{E}\big[H_{\rho(T,x)}\big] = \lim_{n\to\infty} \mathbb{E}\left[\int_0^{\rho(T,x)\wedge\theta_n\wedge n} h(X_s^x)ds\right] = \lim_{n\to\infty} \frac{1}{c}\mathbb{E}\left[\log\Big(g\big(X^x_{\rho(T,x)\wedge\theta_n\wedge n}\big)\Big) - \log\big(g(x)\big)\right].$$

For all examples of Table II.2 it holds that $g(x) \le 1$. Furthermore, we have

$$\log\Big(g\big(X^x_{\rho(T,x)\wedge\theta_n\wedge n}\big)\Big) \ge cT + \log\big(g(x)\big), \quad n \in \mathbb{N},$$

by the definition of $\rho(T,x)$. Hence, dominated convergence implies that

$$\mathbb{E}\left[H^x_{\rho(T,x)}\right] = \lim_{n\to\infty} \frac{1}{c}\mathbb{E}\left[\log\Big(g\big(X^x_{\rho(T,x)\wedge\theta_n\wedge n}\big)\Big) - \log\big(g(x)\big)\right] = T.$$

Therefore, we have $\alpha^* \in \mathcal{A}$ and $\rho(T,x) < \infty$, $\mathbb{P}$-a.s., by Assumption (**A**). For all examples of Table II.2 the function $u(T,x) = -e^{cT}g(x)$, $(T,x) \in [0,\infty) \times \mathbb{R}$, with $c < 0$ is the value function of the optimal control problem $(\mathcal{U})$ with constraint function h and payoff function $-g$ by Theorem II.4.1. An optimal control is given by α^*. The corresponding optimal stopping time for $u(T,x)$ is given by $\rho(T,x)$. Since u is strictly increasing in T we conclude from Theorem II.4.1, 2b) that also the value function V of the control problem $(\mathcal{V})$ is given by u with optimal control (α^*, T). According to Proposition II.2.4 the corresponding optimal stopping time is given by $\rho(T,x)$.

Remark II.5.2. Note that if $a = c$, then $k(x)$ is a solution to (II.5.4) if and only if $\exp\big(k(x)\big)$ is a solution to (II.5.7).

Table II.1: **Additive Structure:** Examples for the functions k and g for different b, σ and h using the ansatz $u(T,x) = g\big(aT + k(x)\big)$, $a \in \mathbb{R}\backslash\{0\}$, where u solves (II.3.3).

$\sigma(x)$	$b(x)$	$h(x)$	$k(x)$	$g(x)$	$\rho(T,x)$
1	0	1	$ax^2 + \operatorname{sgn}(a)$	$(\operatorname{sgn}(a)x)^s$, $s \in \left(0, \frac{1}{2}\right]$	$\inf\left\{t \in [0,\infty) \colon \lvert X_t^x\rvert \geq \sqrt{T + x^2}\right\}$
				$\log(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \lvert X_t^x\rvert \geq \sqrt{T + x^2}\right\}$
				$\operatorname{arcosh}(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \lvert X_t^x\rvert \geq \sqrt{T + x^2}\right\}$
				$\frac{1}{x}$, if $a < 0$	$\inf\left\{t \in [0,\infty) \colon \lvert X_t^x\rvert \geq \sqrt{T + x^2}\right\}$
1	0	$b_1 + b_2x^2$,	$\frac{ab_2}{6}x^4 + ab_1x^2 + \operatorname{sgn}(a)$	$(\operatorname{sgn}(a)x)^s$, $s \in \left(0, \frac{1}{4}\right]$	$\inf\left\{t \in [0,\infty) \colon \left(X_t^x\right)^2 \geq \sqrt{\frac{6T}{b_2} + \left(x^2 + \frac{3b_1}{b_2}\right)^2} - \frac{3b_1}{b_2}\right\}$
		$b_1, b_2 > 0$		$\log(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \left(X_t^x\right)^2 \geq \sqrt{\frac{6T}{b_2} + \left(x^2 + \frac{3b_1}{b_2}\right)^2} - \frac{3b_1}{b_2}\right\}$
1	0	$\cosh(x)$	$2a\cosh(x) + \operatorname{sgn}(a)$	$\operatorname{arcosh}(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \lvert X_t^x\rvert \geq \operatorname{arcosh}\left(T + \cosh(x)\right)\right\}$
1	1	1	$a\left(x + \exp(2-2x)\right) + \operatorname{sgn}(a)$	$\log(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \frac{1}{a}k(X_t^x) \geq T + \frac{1}{a}k(x)\right\}$
1	1	$1 + x^2$	$\frac{a}{12}\left(9e^{2-2x} + 4x^3 - 6x^2 + 18x\right) + \operatorname{sgn}(a)$	$\log(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \frac{1}{a}k(X_t^x) \geq T + \frac{1}{a}k(x)\right\}$
1	1	$\cosh(x)$	$\frac{a}{6}e^{-1-2x}(e - e^x)^2\left(3 + e^2 + 2e^{x+1}\right) + \operatorname{sgn}(a)$	$\operatorname{arcosh}(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \frac{1}{a}k(X_t^x) \geq T + \frac{1}{a}k(x)\right\}$
$\sigma(x) \geq \delta > 0$	0	$\sigma^2(x)$	$a(x-1)^2 + \operatorname{sgn}(a)$	$(\operatorname{sgn}(a)x)^s$, $s \in \left(0, \frac{1}{2}\right]$	$\inf\left\{t \in [0,\infty) \colon \left(X_t^x - 1\right)^2 \geq T + (x-1)^2\right\}$
Lipschitz				$\log(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \left(X_t^x - 1\right)^2 \geq T + (x-1)^2\right\}$
				$\operatorname{arcosh}(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \left(X_t^x - 1\right)^2 \geq T + (x-1)^2\right\}$
$\frac{1}{\sqrt{1+x^2}}$	0	1	$\frac{a}{6}x^4 + ax^2 + \operatorname{sgn}(a)$	$(\operatorname{sgn}(a)x)^s$, $s \in \left(0, \frac{1}{4}\right]$	$\inf\left\{t \in [0,\infty) \colon (X_t^x)^2 \geq \sqrt{6T + (x^2+3)^2} - 3\right\}$
$\frac{1}{\sqrt{1+x^2}}$	0	$1 + x^2$	$\frac{a}{15}x^6 + \frac{a}{3}x^4 + ax^2 + \operatorname{sgn}(a)$	$\log(\operatorname{sgn}(a)x)$	$\inf\{t \in [0,\infty) \colon \lvert X_t^x\rvert = z^*\}$, where z^* is the unique positive solution of $k(z) = aT + k(x)$
$\frac{1}{\sqrt{1+x^2}}$	0	$\sqrt{1+x^2}$	$\frac{a}{20}\left((9+2x^2)x^2\sqrt{1+x^2} - 8\sqrt{1+x^2}\right) + \frac{3a}{4}x\operatorname{arsinh}(x) + \operatorname{sgn}(a) + \frac{2a}{5}$	$\log(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \frac{1}{a}k(X_t^x) \geq T + \frac{1}{a}k(x)\right\}$
$\frac{1}{\sqrt{1+x^2}}$	0	$\cosh(x)$	$2a(7+x^2)\cosh(x) - 8ax\sinh(x) + \operatorname{sgn}(a)$	$\operatorname{arcosh}(\operatorname{sgn}(a)x)$	$\inf\left\{t \in [0,\infty) \colon \frac{1}{a}k(X_t^x) \geq T + \frac{1}{a}k(x)\right\}$

Table II.2: **Multiplicative Structure:** Examples for the function g for different b, σ and h using the ansatz $u(T,x) = -e^{cT}\, g(x)$, $c < 0$, where u is a solution to (II.3.3)

$\sigma(x)$	$b(x)$	$h(x)$	$g(x) = -f(x)$	$\rho(T,x)$
1	0	$b_1 + b_2 x^2$,	$\exp\left(\frac{b_2 c}{6} x^4 + b_1 c x^2\right)$	$\inf\left\{t \in [0,\infty) \colon \left(X_t^x\right)^2 \geq \sqrt{\frac{6T}{b_2} + \left(x^2 + \frac{3b_1}{b_2}\right)^2} - \frac{3b_1}{b_2}\right\}$, $b_2 > 0$,
1	0	$b_1 > 0, b_2 \geq 0$		$\inf\left\{t \in [0,\infty) \colon \left(X_t^x\right)^2 \geq \frac{T}{b_1} + x^2\right\}$, $b_2 = 0$
1	0	$\cosh(x)$	$\exp\left(2c\cosh(x)\right)$	$\inf\left\{t \in [0,\infty) \colon \lvert X_t^x\rvert \geq \operatorname{arcosh}\left(\frac{T}{2} + \cosh(x)\right)\right\}$
1	0	$n''(x) \geq \delta > 0$, $n \in \mathcal{C}^2(\mathbb{R})$, $n(x) \geq K$, $K \in \mathbb{R}$	$\exp\left(2cn(x)\right)$	$\inf\left\{t \in [0,\infty) \colon n\left(X_t^x\right) \geq \frac{T}{2} + n(x)\right\}$
1	1	1	$\exp\left(cx + \frac{c}{2} e^{2-2x}\right)$	$\inf\left\{t \in [0,\infty) \colon 2X_t^x + e^{2-2X_t^x} \geq 2T + 2x + e^{2-2x}\right\}$
1	1	$1 + x^2$	$\exp\left(\frac{c}{12}\left(9e^{2-2x} + 4x^3 - 6x^2 + 18x\right)\right)$	$\inf\left\{t \in [0,\infty) \colon \log\left(g(X_t^x)\right) \leq cT + \log\left(g(x)\right)\right\}$
1	1	$\cosh(x)$	$\exp\left(\frac{c}{6} e^{-1-2x}(e - e^x)^2(3 + e^2 + 2e^{1+x})\right)$	$\inf\left\{t \in [0,\infty) \colon \log\left(g(X_t^x)\right) \leq cT + \log\left(g(x)\right)\right\}$
$s > 0$	$b_1(b_2 - x)$, $b_1 > 0, b_2 \in \mathbb{R}$	$\exp\left(\frac{b_1}{s^2}(x - b_2)^2\right)$	$\exp\left(\frac{c}{b_1} h(x)\right)$	$\inf\left\{t \in [0,\infty) \colon (X_t^x - b_2)^2 \geq \frac{s^2}{b_1} \log\left(b_1 T + h(x)\right)\right\}$
$\sigma(x) \geq \delta > 0$ Lipschitz	0	$\sigma^2(x)$	$\exp\left(cx^2\right)$	$\inf\left\{t \in [0,\infty) \colon (X_t^x)^2 \geq T + x^2\right\}$
$\frac{1}{\sqrt{1+x^2}}$	0	1	$\exp\left(\frac{c}{6} x^4 + cx^2\right)$	$\inf\left\{t \in [0,\infty) \colon (X_t^x)^2 \geq \sqrt{6T + (x^2+3)^2} - 3\right\}$
$\frac{1}{\sqrt{1+x^2}}$	0	$1 + x^2$	$\exp\left(\frac{c}{15} x^6 + \frac{c}{3} x^4 + cx^2\right)$	$\inf\{t \in [0,\infty) \colon \lvert X_t^x\rvert = y(T,x)\}$, where $y(T,x)$ is the unique positive solution of $\log\left(g(y)\right) = cT + \log\left(g(x)\right)$
$\frac{1}{\sqrt{1+x^2}}$	0	$\cosh(x)$	$\exp\left(2c(7 + x^2)\cosh(x) - 8cx\sinh(x)\right)$	$\inf\{t \in [0,\infty) \colon \lvert X_t^x\rvert = y(T,x)\}$, where $y(T,x)$ is the unique positive solution of $\log\left(g(y)\right) = cT + \log\left(g(x)\right)$

II.6 The Lagrangian Dual Problem

In this section we briefly compare our solution method with a Lagrange approach for solving the stopping problem (II.1.2).

We first show that the concave conjugate of $V(T,x)$, considered as a function in T, is the value function of an unconstrained stopping problem with infinite time horizon. To this end we define $\mathcal{T} = \bigcup_{T\in[0,\infty)} \mathcal{T}(T)$.

Proposition II.6.1 (cf. [39])**.** *Let $w\colon [0,\infty)\times\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ denote the function*

$$w(\lambda,x) = \sup_{\tau\in\mathcal{T}} \mathbb{E}\left[f(X^x_\tau) - \lambda H^x_\tau\right] \tag{II.6.1}$$

and let $V^*\colon [0,\infty)\times\mathbb{R}^n \to \mathbb{R}\cup\{-\infty\}$ be the concave conjugate of V with respect to the first argument, i.e.

$$V^*(\lambda,x) = \inf_{T\in[0,\infty)} \big(T\lambda - V(T,x)\big).$$

Then it holds that $V^* = -w$.

Proof. We have for all $(\lambda,x)\in[0,\infty)\times\mathbb{R}^n$

$$-V^*(\lambda,x) = \sup_{T\in[0,\infty)} \left(V(T,x) - T\lambda\right) \le \sup_{T\in[0,\infty)} \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}\left[f(X^x_\tau) - \lambda H^x_\tau\right] = w(\lambda,x).$$

Moreover, for all $x\in\mathbb{R}^n$ and stopping times $\tau\in\mathcal{T}$ it holds that $\mathbb{E}[f(X^x_\tau)] \le V\left(\mathbb{E}\left[H^x_\tau\right],x\right)$. This implies for all $(\lambda,x)\in[0,\infty)\times\mathbb{R}^n$ that

$$w(\lambda,x) \le \sup_{\tau\in\mathcal{T}} \mathbb{E}\left[V\left(\mathbb{E}\left[H^x_\tau\right],x\right) - \lambda H^x_\tau\right] \le \sup_{T\in[0,\infty)} \left(V(T,x) - \lambda T\right) = -V^*(\lambda,x).$$

This completes the proof of the proposition. □

Remark II.6.2. The preceding proof relies on arguments used in the introduction of [39], where a discrete-time variant of the constrained optimal stopping problem (I.1) is analyzed. Using randomized stopping times, Kennedy shows in Section 3 of [39] that $V^{**} = V$, where $V^{**} = (V^*)^*$. Thus, the identity $V^* = -w$ implies that $V = (-w)^*$ and one can recover V from the unconstrained stopping problems (II.6.1).

For every $\lambda\in[0,\infty)$ the function $\mathbb{R}^n \ni x \mapsto w(\lambda,x) \in \mathbb{R}\cup\{+\infty\}$ is the value function of a stopping problem. One way to tackle this problem is the classical PDE approach. The associated dynamic programming equation takes for every $\lambda\in[0,\infty)$ the variational form (cf. eg. [55, Chapter IV Section 7 and 8])

$$\min[-\mathcal{L}\widetilde{w}(\lambda,x) + \lambda h(x), \widetilde{w}(\lambda,x) - f(x)] = 0 \quad \text{for all } x\in\mathbb{R}^n. \tag{II.6.2}$$

Note, however, that in general the solution of this equation is not unique. This can be seen as follows. Suppose that $(X^x_t)_{t\in[0,\infty)}$ is a Brownian motion starting in $x\in\mathbb{R}^n$ and let $h(y)=1$ for all $y\in\mathbb{R}^n$. Furthermore, assume that $f\colon\mathbb{R}^n\to\mathbb{R}$ is bounded from above by some constant $c\in\mathbb{R}$. Then the functions $\widetilde{w}^r\colon[0,\infty)\times\mathbb{R}^n\to\mathbb{R}$, $\widetilde{w}^r(\lambda,x) = \frac{\lambda}{d}\|x\|^2 + r$, $r\ge c$ all satisfy Equation (II.6.2). Observe that if V satisfies the duality relation $V = V^{**}$, then the initial condition $V(0,\cdot) = f$ translates to the identity $f(x) = V(0,x) = \inf_{\lambda\ge 0} -V^*(\lambda,x)$ for all $x\in\mathbb{R}^n$. This suggests to impose the further condition $\inf_{\lambda\ge0}\widetilde{w}(\lambda,x) = f(x)$ on the solutions of the PDE (II.6.2). This might lead to uniqueness, but renders the computation of $\widetilde{w}$ more difficult.

If it is still possible to determine w and if $V = V^{**}$, i.e. V is concave in T, then it follows from Proposition II.6.1 that one can recover V from w via the identity $V = (-w)^*$. Moreover, if $\tau(\lambda,x)$ is optimal in the stopping problem (II.6.1) for $(\lambda,x)\in[0,\infty)\times\mathbb{R}^n$ and satisfies the constraint $\mathbb{E}\left[H^x_{\tau(\lambda,x)}\right] = T$ for some $T \in [0,\infty)$, then $\tau(\lambda,x)$ is also optimal in the problem $V(T,x) = \sup_{\tau\in\mathcal{S}(T)} \mathbb{E}[f(X^x_\tau)]$. Indeed, for every $\tau \in \mathcal{S}(T)$ it holds that

$$\mathbb{E}\left[f\big(X^x_{\tau(\lambda,x)}\big)\right] = w(\lambda,x) + \lambda T \geq \mathbb{E}\left[f(X^x_\tau) - \lambda H^x_\tau\right] + \lambda T \geq \mathbb{E}\left[f(X^x_\tau)\right].$$

Under the mild conditions that f is upper semi-continuous and $\widetilde{w}$ is lower semi-continuous, for a fixed $(\lambda,x) \in [0,\infty)\times\mathbb{R}^n$ an optimal stopping time in (II.6.1) is the first hitting time of the stopping region, i.e. $\tau(\lambda,x) = \inf\{s\in[0,\infty)\colon \widetilde{w}(\lambda,X^x_s) \leq f(X^x_s)\}$ (see Corollary 2.9, Chapter I in [55]). However, it may happen that there exist multiple optimal stopping times for the dual problem (II.6.1). In this case one has to identify the stopping time matching the expectation constraint $\mathbb{E}\big[H^x_{\tau(\lambda,x)}\big] = T$ among all optimal stopping times. To illustrate this fact we revisit Example II.4.9, where optimal stopping times for (II.6.1) also include stopping times that are not hitting times of two points.

Example II.6.3 (cf. Example II.4.9)**.** Let $d = n = 1$, $f\colon \mathbb{R}\to\mathbb{R}, f(y) = y^2\mathbb{1}_{\{|y|\geq 1\}}$, $h\colon \mathbb{R}\to\mathbb{R}, h(y) = 1$ and let X be a one-dimensional Brownian motion which starts in 0. Using (II.4.8) it is straightforward to show that the dual problem satisfies

$$w(\lambda,0) = \sup_{\mathbb{E}[\tau]<\infty} \mathbb{E}\left[(\mathbb{1}_{\{|W_\tau|\geq 1\}} - \lambda)W_\tau^2\right] = \begin{cases}\infty, & \text{if } \lambda < 1,\\ 0, & \text{if } \lambda \geq 1.\end{cases}$$

This implies that $(-w)^*(T,0) = T$ for all $T\in[0,\infty)$. Therefore, it follows from Example II.4.9 that $V(T,0) = (-w)^*(T,0)$ for all $T\in[0,\infty)$. For $\lambda < 1$ there exist no optimal stopping times and for $\lambda \geq 1$ stopping immediately $\tau = 0$ is optimal. But for $\lambda = 1$ all integrable stopping times τ_a that embed the distribution $\frac{a}{2}\delta_{-1} + (1-a)\delta_0 + \frac{a}{2}\delta_1$, $a\in[0,1]$ into W are also optimal. It holds that $\mathbb{E}[\tau_a] = a$. Thus, τ_T is optimal in the primal problem $V(T,0) = \sup_{\tau\in\mathcal{S}(T)}\mathbb{E}[f(W_\tau)]$ for all $T\in[0,1]$.

II.7 Examples with no Optimal Stopping Time

In this section we present examples where no optimal stopping time exists for U, at least for some starting points. In the examples with smooth value function U, the function U does *not* solve the dynamic programming equation (II.3.3).

II.7.1 An Optimal Stopping Time does not exist for $U(T,0)$

We now present an example where no optimal stopping time exists for $U(T,0)$, but for $U(T,x)$ if $x\neq 0$. Moreover, the value function does not solve the PDE (II.3.3) in $(T,0)$. In addition, V coincides with U and it is a classical solution to (II.4.6).

Let $f(y) = e^{-|y|}$, $h(y) = 1$, $y\in\mathbb{R}$, and let $X^x_t = x + W_t$, $t\in[0,\infty)$, be a Brownian motion starting in $x\in\mathbb{R}$. Then,

$$U(T,x) = \begin{cases}1, & \text{if } x = 0,\\ \frac{1}{T+x^2}\left(T + x^2 e^{-|x|-\frac{T}{|x|}}\right), & \text{if } x\neq 0,\end{cases} \tag{II.7.1}$$

$U\in\mathcal{C}^2\big((0,\infty)\times\mathbb{R}\big)$, but there exists no optimal stopping time for $U(T,0)$, $T\in(0,\infty)$. In addition, U is a solution to (II.3.3) on $(0,\infty)\times\mathbb{R}\backslash\{0\}$ but only a strict supersolution on $(0,\infty)\times\{0\}$:

$$h(x)U_T(T,x) - \frac{1}{2}U_{xx}(T,x) + \frac{U^2_{Tx}(T,x)}{2U_{TT}(T,x)} = \begin{cases}0, & \text{if } x\neq 0,\\ \frac{1}{T}, & \text{if } x = 0.\end{cases}$$

We first show that $U(T,0)=1$ and that there does not exist an optimal stopping time for $U(T,0)$. To see this, let $T\in(0,\infty)$ and notice that the first exit time $\rho(T,\varepsilon)$ of $(-\varepsilon,T/\varepsilon)$, $\varepsilon>0$, has expectation T. Thus,

$$U(T,0)\geq \sup_{\varepsilon>0}\mathbb{E}\left[f\left(X^0_{\rho(T,\varepsilon)}\right)\right]=\sup_{\varepsilon>0}\left\{\frac{Te^{-\varepsilon}}{T+\varepsilon^2}+\frac{\varepsilon^2 e^{-\frac{T}{\varepsilon}}}{T+\varepsilon^2}\right\}=1.$$

Since f is bounded above by 1, we conclude that $U(T,0)=1$. Now assume that there exists an optimal stopping time τ^* for $U(T,0)$ with $T\in(0,\infty)$, i.e. $\mathbb{E}[f(X^0_{\tau^*})]=1$ and $\mathbb{E}[\tau^*]=T$. This implies $X^0_{\tau^*}=0$, $\mathbb{P}$-a.s., but this is not possible if τ^* is integrable with $\mathbb{E}[\tau^*]>0$. Hence, an optimal stopping time for $U(T,0)$, $T\in(0,\infty)$, does not exist.

Now let $x\neq 0$. For $T\in(0,\infty)$ the reward of the first exit time $\rho_x(T)$ of $\left(0,x+\frac{T}{x}\right)$ if $x>0$, and $\left(x+\frac{T}{x},0\right)$ if $x<0$, is given by

$$\mathbb{E}\left[f\left(X^x_{\rho_x(T)}\right)\right]=\frac{1}{T+x^2}\left(T+x^2e^{-|x|-\frac{T}{|x|}}\right)=: u(T,x).$$

It holds true that $u\in\mathcal{C}^2\big((0,\infty)\times\mathbb{R}\big)$ and u is a solution to (II.3.3) on $(0,\infty)\times\mathbb{R}\backslash\{0\}$. But for $T\in(0,\infty)$ we have

$$h(x)u_T(T,0)-\frac{1}{2}u_{xx}(T,0)+\frac{1}{2}\frac{u^2_{Tx}(T,0)}{u_{TT}(T,0)}=0-\frac{1}{2}\left(-\frac{2}{T}\right)+\frac{1}{2}0=\frac{1}{T}.$$

Hence, u is a supersolution to (II.3.3). Moreover, $u(0,.)=f$, u is concave with $u_{TT}(T,x)<0$ for $x\neq 0$ and $u_{Tx}(T,x)=0$ on $\{(T,x)\in(0,\infty)\times\mathbb{R}\colon u_{TT}(T,x)=0\}=(0,\infty)\times\{0\}$. Since $|u|$ is bounded above by 1, u satisfies the growth condition (II.4.1) with $p=1$. In addition, $\mathcal{X}^1_\tau$ is uniformly integrable for all $\tau\in\mathcal{S}(T)$ by Remark II.4.2. Hence, Theorem II.4.1 implies that u dominates the value function and, therefore, $U=u$.

Remember that the quotient U^2_{Tx}/U_{TT} is set to 0 if the nominator and the denominator equal 0. Here we have

$$\frac{U^2_{Tx}(T,x)}{U_{TT}(T,x)}\xrightarrow[x\to 0]{}-\frac{2}{T}\neq 0=\frac{U^2_{Tx}(T,0)}{U_{TT}(T,0)}.$$

Hence, $(T,x)\mapsto U^2_{Tx}(T,x)/U_{TT}(T,x)$ is not continuous on $(0,\infty)\times\{0\}$ and, therefore, U cannot be a solution to (II.3.3) by Remark II.3.5b).

Bayraktar and Yao [7] characterize the value function of an optimal stopping problem with expectation constraint as a viscosity supersolution of the HJB equation and as a viscosity subsolution to the HJB equation, in which the Hamiltonian is replaced by its upper semi-continuous envelope, see Remark II.3.3.

In this example we obtain a similar result. In the HJB equation (II.3.3) the Hamiltonian $\mathcal{H}$ is given by $\mathcal{H}v(T,x)=-\left|\sigma^\top\cdot\nabla_x v_T(T,x)\right|^2/2v_{TT}(T,x)$. For the value function U given in (II.7.1) the upper semi-continuous envelope $(\mathcal{H}U)^*$ of $\mathcal{H}U$ satisfies

$$(\mathcal{H}U)^*(T,x)=\begin{cases}-\frac{U^2_{Tx}(T,x)}{2U_{TT}(T,x)}, & \text{if } x\neq 0,\\ \frac{1}{T} & \text{if } x=0.\end{cases}$$

Notice that the function U solves for all $(T,x)\in(0,\infty)\times\mathbb{R}$

$$h(x)U_T(T,x)-\frac{1}{2}U_{xx}(T,x)-(\mathcal{H}U)^*(T,x)=0.$$

It is worth mentioning that $\tau^*=0$ is an optimal stopping time for $V(T,0)$. Moreover, $V(T,x)=U(T,x)$, because $U_T(T,x)>0$ for $x\neq 0$. In particular, V is a solution to (II.4.6) on the whole domain $(0,\infty)\times\mathbb{R}$.

II.7.2 Maximizing a Concave Payoff Function

We now show that if X is a Brownian motion and f a concave payoff function, then the value function U is only a strict supersolution to (II.3.3). Moreover, U is independent of T and coincides with f.

Let $d = n = 1$, $h(y) = 1, y \in \mathbb{R}$, and $X_t^x = x + W_t$, $t \in [0, \infty)$, $x \in \mathbb{R}$. For every concave and continuous f with subquadratic growth, i.e. $\lim_{y \to \pm\infty} |f(y)|/y^2 = 0$, the value functions U and V are given by

$$U(T, x) = V(T, x) = f(x)$$

for all $(T, x) \in [0, \infty) \times \mathbb{R}$. On the one hand the function $f(x)$ dominates $U(T, x)$ and $V(T, x)$ by Jensen's inequality. To prove the reverse inequality, we define for every $(T, x) \in (0, \infty) \times \mathbb{R}$ and $\varepsilon > 0$ the stopping times

$$\rho_x(\varepsilon, T) = \inf\left\{t \in [0, \infty) \colon X_t^x \notin \left(x - \varepsilon, x + \frac{T}{\varepsilon}\right)\right\}.$$

Notice that $\mathbb{E}[\rho_x(\varepsilon, T)] = T$, and hence we have

$$\begin{aligned}
U(T, x) &\geq \sup_{\varepsilon > 0} \mathbb{E}\left[f\left(X^x_{\rho_x(\varepsilon, T)}\right)\right] \\
&= \sup_{\varepsilon > 0}\left\{f\left(x + \frac{T}{\varepsilon}\right)\frac{\varepsilon^2}{T + \varepsilon^2} + f(x - \varepsilon)\frac{T}{T + \varepsilon^2}\right\} \\
&\geq \lim_{\varepsilon \to 0}\left\{f\left(x + \frac{T}{\varepsilon}\right)\frac{\varepsilon^2}{T + \varepsilon^2} + f(x - \varepsilon)\frac{T}{T + \varepsilon^2}\right\} \\
&= f(x).
\end{aligned}$$

If, in addition, f is twice continuously differentiable and there exists $y \in \mathbb{R}$ with $f''(y) < 0$, then U is not a solution to (II.3.3) on $(0, \infty) \times \{y \in \mathbb{R} \colon f''(y) < 0\}$. Notice, however, V is a solution to (II.4.6) on $(0, \infty) \times \mathbb{R}^n$.

II.7.3 The Value Function is linear in T

In the previous two examples the value function U does not solve the PDE (II.3.3), but V is a solution to (II.4.6). We now present an example where both U and V do not solve the corresponding dynamic programming equation.

Let $f(y) = y^2 - \sqrt{1 + y^2}$, $h(y) = 1, y \in \mathbb{R}$, and let $X_t^x, t \in [0, \infty)$, be a one-dimensional Brownian motion with $X_0^x = x \in \mathbb{R}$. For every stopping time $\tau \in \mathcal{S}(T)$, $T \in (0, \infty)$, we have

$$\mathbb{E}[f(X_\tau^x)] = x^2 + \mathbb{E}[\tau] + \mathbb{E}\left[-\sqrt{1 + (X_\tau^x)^2}\right] \leq x^2 + T - \sqrt{1 + x^2} = T + f(x)$$

by (II.4.8) and Jensen's inequality. Hence, $U(T, x) \leq V(T, x) \leq T + f(x)$.

As in Example II.7.2 we define for every $(T, x) \in (0, \infty) \times \mathbb{R}$ and $\varepsilon > 0$ the stopping times

$$\rho_x(\varepsilon, T) = \inf\left\{t \in [0, \infty) \colon X_t^x \notin \left(x - \varepsilon, x + \frac{T}{\varepsilon}\right)\right\}.$$

Then it follows that

$$\begin{aligned}
U(T, x) \geq \sup_{\varepsilon > 0} \mathbb{E}\left[f\left(X^x_{\rho_x(\varepsilon, T)}\right)\right] &= \sup_{\varepsilon > 0}\left\{\frac{\varepsilon^2}{T + \varepsilon^2} f\left(x + \frac{T}{\varepsilon}\right) + \frac{T}{T + \varepsilon^2} f(x - \varepsilon)\right\} \\
= T + x^2 + \sup_{\varepsilon > 0}&\left\{-\sqrt{1 + \left(x + \frac{T}{\varepsilon}\right)^2}\frac{\varepsilon^2}{T + \varepsilon^2} - \sqrt{1 + (x - \varepsilon)^2}\frac{T}{T + \varepsilon^2}\right\} \\
\geq T + f(x).&
\end{aligned}$$

Therefore, $V(T,x) = U(T,x) = T + f(x)$. U is not a solution to (II.3.3), because

$$h(x)U_T(T,x) - \mathcal{L}U(T,x) + \frac{U_{Tx}^2(T,x)}{2U_{TT}(T,x)} = \frac{1}{2(1+|x|^2)^{3/2}} > 0.$$

Moreover, V does not solve the PDE (II.4.6). Indeed,

$$\min\left\{V_T(T,x), h(x)V_T(T,x) - \mathcal{L}V(T,x) + \frac{V_{Tx}^2(T,x)}{2V_{TT}(T,x)}\right\} = \min\left\{1, \frac{1}{2(1+|x|^2)^{3/2}}\right\} > 0.$$

III. 3 Points Suffice

In this chapter we focus on optimally stopping a one-dimensional regular continuous strong Markov process Y with stopping times τ satisfying the expectation constraint $\mathbb{E}[\tau] \leq T$, $T \in [0, \infty)$. Observe that on the one hand we restrict the cost constraints from Chapter II to an expectation constraint and that we only study one-dimensional processes. But on the other hand, we allow for processes that cannot be described by a solution of a time-homogeneous stochastic differential equation driven by a Brownian motion.

We show that in an optimal stopping problem with expectation constraint we can confine to stopping times τ such that the law of Y_τ is a weighted sum of at most 3 Dirac measures. Consecutive exit times from intervals can be used to embed a discrete probability measure with at most 3 mass points in Y.

Furthermore, a reduction to weighted sums of 2 Dirac measures and hence, to exit times of intervals is in general not possible. Consequently, we cannot identify a deterministic stopping and continuation region as in stopping problems with a sharp bound on the stopping time or with infinite time horizon and discounting.

We first rewrite the stopping problem as a linear optimization problem over a set of probability measures in Section III.2. By using results on the Skorokhod embedding problem we identify the set $\mathcal{A}(T)$ of distributions of Y_τ, where τ is a stopping times having expectation smaller than or equal to T, see [2] and [32]. The balayage method of Chacon and Walsh (see [17]) allows to construct for every $\mu \in \mathcal{A}(T)$ an approximating sequence of probability measures that are weighted sums of finitely many Dirac measures (Section III.3). Thus, we can restrict the set of probability measures $\mathcal{A}(T)$ in the optimization problem to discrete measures with finitely many points of mass (Section III.4.1). By a result of Hoeffding [33] we can further confine to the set $\mathcal{A}_3(T)$ of probability measures in $\mathcal{A}(T)$ that are weighted sums of at most 3 Dirac measures (Section III.4.2).

To extend the results from Section III.4 to more general payoff functions in Section III.5, we take advantage of the linearity of the measure optimization problem. As in standard linear problems, the value of the measure optimization problem over $\mathcal{A}(T)$ is attained by extreme points of $\mathcal{A}(T)$. It turns out that the extreme points are contained in the set of probability measures that can be written as a weighted sum of at most 3 Dirac measures.

In particular, the question arises whether an optimal stopping time exists in the constrained problem. Hence, we provide sufficient conditions for the existence of an optimal measure $\mu^* \in \mathcal{A}_3(T)$ in the measure optimization problem under mild conditions on the payoff function in Section III.6. We use the Balayage method to construct an embedding stopping time τ^* for μ^* in Y. Then τ^* is optimal in the stopping problem with expectation constraint. Throughout this chapter we assume that the process to stop is in natural scale - this is, as explained in Section III.7, not a restriction. Finally, in Section III.8 we compare the Lagrange approach for solving constrained optimal stopping problems to our method of reducing the set of stopping times.

Sections III.1, III.2, III.5, III.6 and III.7 are based on a revised version of [3].

III.1 Stopping after Consecutive Exit Times

In this section we rigorously set the framework for the optimal stopping problem. The process to stop is assumed to be a one-dimensional regular continuous strong Markov process. In the sequel we use the term general diffusion as a synonym for these processes. Let the state space $J \subseteq \mathbb{R}$ be an open, half-open or closed interval and denote by (l,r) the interior of J, where $-\infty \le l < r \le \infty$. Moreover, denote by $\overline{J}$ the closure of J in $\mathbb{R}$. Let $\Omega = \mathcal{C}\big([0,\infty), J\big)$ be the space of all continuous J-valued functions and $(Y_t)_{t\in[0,\infty)}$ be the coordinate process, i.e. $Y_t(\omega) = \omega(t)$, $t \in [0,\infty)$, $\omega \in \Omega$. Let $\mathcal{F}_t^0$ be the σ-algebra generated by $(Y_s)_{s\in[0,t]}$ and $\mathcal{F}^0 := \mathcal{F}_\infty^0 := \bigvee_{t\in[0,\infty)} \mathcal{F}_t^0$. Denote by $(\theta_t)_{t\in[0,\infty)}$ the family of shift operators on Ω defined by $(\theta_t\omega)(s) = \omega(t+s)$, $s \in [0,\infty)$. Let $(\mathbb{P}^y)_{y\in J}$ be a family of probability measures on $(\Omega, \mathcal{F}^0)$ that is a regular diffusion in the sense of [59, Chapter V.45]. In particular, we have $\mathbb{P}^y[Y_0 = y] = 1$ for all $y \in J$. Regularity means that for every $y \in (l,r)$ and $x \in J$ it holds that $\mathbb{P}^y[\tau_x < \infty] > 0$, where $\tau_x = \inf\{t \in [0,\infty) \colon Y_t = x\}$. Here and in the sequel we use the convention that $\inf \emptyset = \infty$.

For a probability measure ν on $\big(J, \mathcal{B}(J)\big)$ let

$$\mathbb{P}^\nu[A] := \int_J \mathbb{P}^y[A]\nu(dy), \qquad A \in \mathcal{F}^0.$$

Let $\mathcal{F}^\nu$ be the completion of $\mathcal{F}^0$ with respect to $\mathbb{P}^\nu$ and set $\mathcal{F}_t^\nu = \sigma(\mathcal{F}_t^0, \mathcal{N})$, $t \in [0,\infty)$, where $\mathcal{N}$ denotes the collection of $\mathbb{P}^\nu$-null sets in $\mathcal{F}^\nu$. One can show that $(\Omega, \mathcal{F}^\nu, (\mathcal{F}_t^\nu), \mathbb{P}^\nu)$ satisfies the usual conditions. We set $\mathcal{F}_t = \bigcap_\nu \mathcal{F}_t^\nu$ and $\mathcal{F} = \bigcap_\nu \mathcal{F}^\nu$. Observe that $(\mathcal{F}_t)$ is right-continuous, but that in general $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}^\nu)$ does not satisfy the usual conditions. The process $(Y_t)_{t\in[0,\infty)}$ fulfills the strong Markov property (cf. Theorem 9.4, Chapter III, in [58]): For any bounded $\mathcal{F}$-measurable mapping η and any finite $(\mathcal{F}_t)$-stopping time τ we have

$$\mathbb{E}^\nu\left[\eta \circ \theta_\tau \,|\, \mathcal{F}_\tau\right] = \mathbb{E}^{Y_\tau}[\eta], \qquad \mathbb{P}^\nu\text{ - a.s.}$$

Let m be the speed measure of the diffusion $(\mathbb{P}^y)_{y\in J}$ on J (see Theorem 3.6, Definition 3.7 and Proposition 3.10 in Chapter VII of [56]). More precisely, m is the unique Radon measure on (l,r) such that for any open subinterval (a,b), $a < b$, with $[a,b] \subseteq J$ it holds that

$$\mathbb{E}^y[\inf\{t \in [0,\infty) \colon Y_t \notin (a,b)\}] = \int_{[a,b]} \frac{\big((s(x)\wedge s(y)) - s(a)\big)\big(s(b) - \big(s(x) \vee s(y)\big)\big)}{s(b) - s(a)}\, m(dx).$$

Here s denotes the scale function of Y, i.e. $s \colon J \to \mathbb{R}$ is strictly increasing and continuous such that for all $a, b, y \in J$ with $a < y < b$ it holds that

$$\mathbb{P}^y[\tau_a < \tau_b] = \frac{s(b) - s(y)}{s(b) - s(a)},$$

see Definition 46.10 in [59], Chapter V. Note that s is unique up to increasing affine transformations. Since Y is regular we have for all $a, b \in (l,r)$ with $a < b$

$$0 < m\big([a,b]\big) < \infty.$$

In Section III.1 – III.6 we assume that the diffusion Y is in natural scale, that is $s(x) = x$. If Y is not in natural scale, then $\big(s(Y_t)\big)_{t\in[0,\infty)}$ is in natural scale, see e.g. Theorem 46.12 in [59]. In Section III.7 below we show how to reduce the stopping problem for general diffusions to the case where the process to stop is in natural scale.

In addition, we assume that if an endpoint c of the interval J is accessible, i.e. $c \in J$, then it is absorbing, that is $\mathbb{P}^c[\tau_x < \infty] = 0$ for all $x \in J\backslash\{c\}$. This implies that Y is a local martingale (see Corollary 46.15 in [59]).

For $y \in (l,r)$ we define $q_y \colon \bar{J} \to [0,\infty]$,

$$q_y(x) = \frac{1}{2} m\big(\{y\}\big)|x - y| + \int_y^x m\big((y,u)\big)\, du, \tag{III.1.1}$$

with the convention that $m\big((y,u)\big) = -m\big((u,y)\big)$ whenever $u < y$. Moreover, we set $q_y(r) := \lim_{x\uparrow r} q_y(x) = \infty$ if $r = \infty$ and $q_y(l) := \lim_{x\downarrow l} q_y(x) = \infty$ if $l = -\infty$. Observe that q_y is convex and continuous on J.

Let $\tau_{l,r} = \inf\{t \in [0,\infty)\colon Y_t \notin (l,r)\}$. One can show that $q_y(Y_t) - (t \wedge \tau_{l,r})$, $t \in [0,\infty)$, is a local martingale with respect to $\mathbb{P}^y$ and $(\mathcal{F}_t)$ (see Theorem 2.1 in [5]). Moreover, the behavior of q_y at l and r determines whether the process attains the boundary points with a positive probability or not.

Lemma III.1.1 (see Theorem 3.3 in [5])**.** *We have $q_y(r) < \infty$ if and only if $r \in J$. Similarly, $q_y(l) < \infty$ if and only if $l \in J$.*

Remark III.1.2. Observe that the case where the process to stop is described by a homogeneous stochastic differential equation (SDE) driven by a Brownian motion W is a special case of the general framework that we set up above. Indeed, let $b, \eta\colon \mathbb{R} \to \mathbb{R}$ be Borel-measurable functions that satisfy $\eta(x) \neq 0$ for all $x \in (l,r)$, $\eta(x) = 0$ for all $x \in \mathbb{R} \setminus (l,r)$ and $\frac{1+|b|}{\eta^2} \in L^1_{loc}\big((l,r)\big)$. Then for all $y \in (l,r)$ the SDE

$$dY_t = b(Y_t)dt + \eta(Y_t)dW_t, \quad Y_0 = y, \tag{III.1.2}$$

possesses a weak solution (Y,W) that is unique in law (see e.g. Theorem 2.11 in [18] or Section 5.5 C in [37]). If $b \equiv 0$, then Y is in natural scale and the results of Section III.1–III.6 apply. In this case the speed measure of Y is given by $m(dx) = \frac{2}{\eta^2(x)}dx$ and for all $y \in (l,r)$ the function q_y satisfies

$$q_y(x) = \int_y^x \int_y^z \frac{2}{\eta^2(u)} du\, dz, \quad x \in \bar{J}.$$

The case where the SDE (III.1.2) contains a non-zero drift component b is a special case of the setting considered in Section III.7.

Let $f\colon J \to \mathbb{R}$ be a Borel-measurable function determining the payoff of the stopping problem. Throughout we make the following assumption on f:

Assumption. For every $y \in (l,r)$ there exists $C(y) \in [0,\infty)$ such that

$$f(x) \geq -C(y)\big(1 + q_y(x)\big), \qquad x \in J. \tag{\textbf{B}}$$

For any $T \in [0,\infty)$, let $\mathcal{S}(T,y)$ be the set of all $(\mathcal{F}_t)$-stopping times τ with $\mathbb{E}^y[\tau] \leq T$.

Remark III.1.3. Assumption (**B**) ensures that the expectation $\mathbb{E}^y[f(Y_\tau)]$ exists for all $y \in (l,r)$, $T \in [0,\infty)$ and $\tau \in \mathcal{S}(T,y)$. Indeed, denote by $\{f(Y_\tau)\}^-$ the negative part of $f(Y_\tau)$. Then for an appropriately chosen localizing sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ for the local martingale $\big(q_y(Y_t) - (t \wedge \tau_{l,r})\big)_{t\in[0,\infty)}$, it holds that

$$\begin{aligned}
\mathbb{E}^y\big[\{f(Y_\tau)\}^-\big] &\leq \mathbb{E}^y\big[C(y)\big(1 + q_y(Y_\tau)\big)\big] = C(y)\left(1 + \mathbb{E}^y\left[\liminf_{n\to\infty} q_y(Y_{\tau\wedge\tau_n\wedge n})\right]\right) \\
&\leq C(y)\left(1 + \liminf_{n\to\infty} \mathbb{E}^y\big[q_y(Y_{\tau\wedge\tau_n\wedge n})\big]\right) = C(y)\big(1 + \liminf_{n\to\infty} \mathbb{E}^y[\tau \wedge \tau_n \wedge n]\big) \\
&\leq C(y)(1+T).
\end{aligned}$$

We consider the problem of finding the stopping time in $\mathcal{S}(T,y)$ that maximizes the expected payoff $\mathbb{E}^y[f(Y_\tau)]$. The value function is defined by

$$V(T,y) = \sup_{\tau\in\mathcal{S}(T,y)} \mathbb{E}^y[f(Y_\tau)] \tag{III.1.3}$$

for $T \in [0,\infty)$ and $y \in J$. Observe that $V(0,y) = f(y)$ for all $y \in J$. Moreover, if $r \in J$, then we have $V(T,r) = f(r)$ for all $T \in [0,\infty)$ and similarly, if $l \in J$, then $V(T,l) = f(l)$, $T \in [0,\infty)$, because we assume that every accessible endpoint is absorbing. Therefore, we assume throughout this chapter that $y \in (l,r)$.

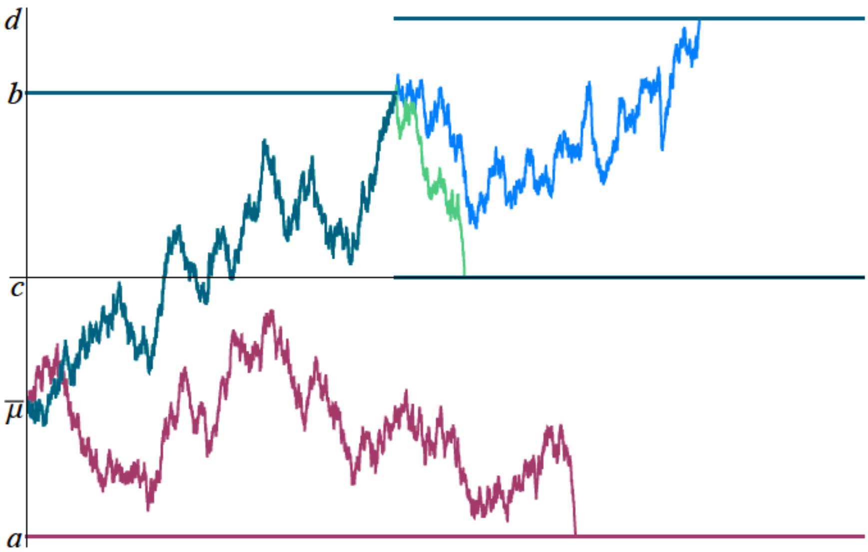


Figure III.1: Three realizations of a Brownian motion starting in $\bar{\mu}$ and stopped at the consecutive exit time $\tau = \tau_{a,b}(0) + \mathbb{1}_{\{Y_{\tau_{a,b}(0)}=b\}} \inf\{t \in [0,\infty)\colon Y_{t+\tau_{a,b}(0)} \in \{c,d\}\}$, where $a=-1$, $b=\frac{5}{7}$, $c=0$, $d=1$. Note that the barriers c and d enter and the barrier b vanishes at the time the process hits b.

Remark III.1.4. If Assumption (**B**) is replaced by the stronger assumption that for every $y \in (l,r)$ there exists $C(y) \in [0,\infty)$ such that

$$|f(x)| \le C(y)\big(1+q_y(x)\big), \qquad x \in J, \tag{III.1.4}$$

then the value function $V(T,y)$ is finite and bounded below by $f(y)$. Indeed, it follows by using similar arguments as in Remark III.1.3 or as in the proof of Proposition II.3.1 in Chapter II that $V(T,y) = \sup_{\tau\in\mathcal{S}(T,y)} \mathbb{E}^y[f(Y_\tau)] \le C(y)(1+T)$.

The following example shows that in general we cannot dispense with condition (III.1.4) if we want to guarantee that V is finite. Recall Example II.3.2: For a Brownian motion $Y=W$ we have $q_0(x)=x^2$. Let $f(x)=|x|^{2+\varepsilon}$, $\varepsilon>0$, be the payoff function. For every $T\in(0,\infty)$ and $a\in(0,\infty)$ the first time $\rho(a,T)$ when W hits a or $-\frac{T}{a}$ has expectation T under $\mathbb{P}^0$. Hence,

$$V(T,0) \ge \sup_{a\in(0,\infty)} \mathbb{E}^0\big[f(W_{\rho(a,T)})\big] = \sup_{a\in(0,\infty)} \left\{a^{2+\varepsilon}\frac{T}{a^2+T} + \frac{a^2}{a^{2+\varepsilon}}\frac{T^{2+\varepsilon}}{a^2+T}\right\} = \infty.$$

For stopping problems *without* an expectation constraint an optimal stopping time is given by the exit time of the continuation region (see Corollary 2.9, Chapter I in [55]). In particular, for solving unconstrained stopping problems it is enough to consider exit times from intervals. For constrained stopping problems a reduction to simple exit times is not possible. We show, however, that it is enough to consider at most three consecutive exit times.

To give a precise statement, we denote for $a,b\in\mathbb{R}$ with $a\le b$ the first hitting time of a by $\tau_a = \inf\{t\in[0,\infty)\colon Y_t=a\}$ and the first exit time from the interval (a,b) after time $r\ge 0$ by $\tau_{a,b}(r)=\inf\{t\ge r\colon Y_t\notin(a,b)\}$. Observe that $\tau_{a,b}(r)=r+\tau_{a,b}(0)\circ\theta_r$. Moreover, we write $\mathcal{S}_3(T,y)$ for the collection of stopping times $\tau\in\mathcal{S}(T,y)$ for which there exist $p_1,p_2,p_3\in[0,1]$ with $p_1+p_2+p_3=1$ and $a,c,d\in\mathbb{R}$ with $a\le c\le d$ such that

$$\tau = \tau_{a,b}(\tau_{\bar{\mu}}) + \mathbb{1}_{\{Y_{\tau_{a,b}(\tau_{\bar{\mu}})}=b\}} \inf\{t\in[0,\infty)\colon Y_{t+\tau_{a,b}(\tau_{\bar{\mu}})}\in\{c,d\}\},$$

where $\bar{\mu}=p_1a+p_2c+p_3d$ and $b=\frac{p_2c+p_3d}{1-p_1}$ if $p_1<1$ and $b=a$ if $p_1=1$. Notice that $b\in\big(\max\{c,\bar{\mu}\},d\big)$ if $p_1<1$. In Figure III.1 three realizations of a Brownian motion starting in $\bar{\mu}$ and stopped at a consecutive exit time are illustrated as well as the fixed barriers a,b,c,d.

One of our main results is that the stopping problem (III.1.3) can be simplified to the set $\mathcal{S}_3(T, y)$.

Theorem III.1.5. *We have*

$$V(T, y) = \sup_{\tau \in \mathcal{S}_3(T,y)} \mathbb{E}^y[f(Y_\tau)]. \tag{III.1.5}$$

We prove Theorem III.1.5 in Section III.4. We do so by reducing problem (III.1.3) to an optimization over a set of probability measures in Sections III.2 and III.3.

Theorem III.1.5 raises the question whether the supremum is attained in $\mathcal{S}_3(T, y)$. In Section III.6 below we provide sufficient conditions guaranteeing the existence of an optimal stopping time in $\mathcal{S}_3(T, y)$.

III.2 Optimal Stopping as a Measure Optimization

In this section we first explain how one can reduce the stopping problem (III.1.3) to a linear optimization problem over a set of probability measures satisfying some integrability constraints.

We denote by $\mathcal{M} = \mathcal{M}(J)$ the set of all probability measures on $\mathbb{R}$ with support in J and by $\mathcal{M}^1$ the set of all measures μ in $\mathcal{M}$ with finite first moment $\bar{\mu} = \int_{\mathbb{R}} x \, \mu(dx)$. For $T \in [0, \infty)$ and $y \in (l, r)$ let $\mathcal{A}(T, y)$ be the set of measures $\mu \in \mathcal{M}^1$ satisfying the following properties:

1. a) If $l > -\infty$, then $\bar{\mu} \leq y$.
   b) If $r < \infty$, then $\bar{\mu} \geq y$.
2. μ integrates q_y such that

$$\int_{\mathbb{R}} q_y(x) \mu(dx) \leq T - H(y, \bar{\mu}), \tag{III.2.1}$$

where $H\colon (l, r) \times J \to [0, \infty]$,

$$H(y, \bar{\mu}) = \begin{cases} (y - \bar{\mu})\big(m\big((y, \infty)\big) + \frac{1}{2} m\big(\{y\}\big)\big), & \bar{\mu} < y, \\ 0, & \bar{\mu} = y, \\ (\bar{\mu} - y)\big(m\big((-\infty, y)\big) + \frac{1}{2} m\big(\{y\}\big)\big), & \bar{\mu} > y. \end{cases}$$

Remark III.2.1. Observe that the following consequences of the definition of $\mathcal{A}(T, y)$ hold true. If there exists $\mu \in \mathcal{A}(T, y)$ such that $\bar{\mu} > y$, then it follows that $l = -\infty$ and that $m\big((-\infty, y)\big) < \infty$. Indeed, the fact that $l = -\infty$ follows directly from Condition 1a) in this case. For the second claim, suppose on the contrary that $m\big((-\infty, y)\big) = \infty$. Then it follows from the definition of H that $H(y, \bar{\mu}) = \infty$ and hence (III.2.1) cannot be satisfied by μ. Consequently, $m\big((-\infty, y)\big) < \infty$. Similarly, it holds that $r = \infty$ and that $m\big((y, \infty)\big) < \infty$ if there exists $\mu \in \mathcal{A}(T, y)$ such that $\bar{\mu} < y$.

Results from [32] on the Skorokhod embedding problem for diffusions (and from [2] for processes described in terms of SDEs) imply that $\mathcal{A}(T, y)$ coincides with the set of probability measures μ that can be embedded into Y under $\mathbb{P}^y$ with stopping times τ satisfying $\mathbb{E}^y[\tau] \leq T$. More precisely, we have the following:

Proposition III.2.2. *Let $\mu \in \mathcal{M}$. There exists a stopping time $\tau \in \mathcal{S}(T, y)$ with $Y_\tau \sim \mu$ under $\mathbb{P}^y$ if and only if $\mu \in \mathcal{A}(T, y)$.*

Proof. Let $\tau \in \mathcal{S}(T, y)$ be an embedding of μ in Y under $\mathbb{P}^y$, i.e. let Y_τ have the distribution μ under $\mathbb{P}^y$. Then [20] and [51] imply that if $l > -\infty$, then $\bar{\mu} \leq y$ and if $r < \infty$, then $\bar{\mu} \geq y$. Thus, we conclude that $\mu \in \mathcal{M}^1$ whenever r or l is finite. Section 3.5 in [32] shows that if $J = (-\infty, \infty)$

and τ is an integrable embedding for μ, then $\mu \in \mathcal{M}^1$. If $\bar{\mu} = y$, then it follows from Theorem 2.4. in [32] that

$$\int_{\mathbb{R}} q_y(x)\mu(dx) \le \mathbb{E}^y[\tau] \le T = T - H(y, \bar{\mu}).$$

If $\bar{\mu} < y$, then we conclude from Theorem 3.6 in [32] that

$$\int_{\mathbb{R}} q_y(x)\mu(dx) + (y - \bar{\mu}) \left(m\big((y,\infty)\big) + \frac{1}{2} m\big(\{y\}\big) \right) \le T,$$

which implies that the second property in the definition of $\mathcal{A}(T, y)$ holds true. If $\bar{\mu} > y$, we again apply Theorem 3.6 in [32] to obtain $\int_{\mathbb{R}} q_y(x)\mu(dx) \le T - H(y, \bar{\mu})$. Hence, $\mu \in \mathcal{A}(T, y)$.

For the reverse direction let $\mu \in \mathcal{A}(T, y)$ and assume first that μ is centered around y. Then μ can be embedded in Y under $\mathbb{P}^y$ by [20] and [51]. It follows from Theorem 3.4 in [32] that there exists a stopping time τ with $Y_\tau \sim \mu$ under $\mathbb{P}^y$ and

$$\mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\mu(dx).$$

The stopping time τ is minimal in the sense that whenever $\rho \le \tau$ is an embedding of μ into Y under $\mathbb{P}^y$, then $\tau = \rho$, $\mathbb{P}^y$-a.s. Since $\mu \in \mathcal{A}(T, y)$ we deduce that $\mathbb{E}^y[\tau] \le T$ and therefore, $\tau \in \mathcal{S}(T, y)$.

Now let $\mu \in \mathcal{A}(T, y)$ with $\bar{\mu} < y$. Then we have $r = \infty$, see Remark III.2.1. Theorem 3.6 in [32] shows the existence of a minimal embedding τ of μ in Y under $\mathbb{P}^y$ with

$$\mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\mu(dx) + (y - \bar{\mu}) \left(m\big((y,\infty)\big) + \frac{1}{2} m\big(\{y\}\big) \right) \le T,$$

where the last inequality follows from the second property of μ. Hence, $\tau \in \mathcal{S}(T, y)$.

Finally, for $\mu \in \mathcal{A}(T, y)$ with $\bar{\mu} > y$, using similar arguments, one can show that there exists a stopping time τ with $Y_\tau \sim \mu$ under $\mathbb{P}^y$ and $\mathbb{E}^y[\tau] \le T$. □

Remark III.2.3. The function q_y appearing in the definition of the set of measures $\mathcal{A}(T, y)$ plays for the Markov process Y the same role than the function $x \mapsto x^2$ plays for the Brownian motion. Indeed, we know that when Y is a Brownian motion starting in $y = 0$ under $\mathbb{P}^0$, we can find an embedding of μ under $\mathbb{P}^0$ with an integrable stopping time if and only if μ is centered and $\int_{\mathbb{R}} x^2 \mu(dx) < \infty$ (see Theorem 1 in [62] and Proposition 17 in [45]). The papers [32] and [2] identify the function q_y as the counterpart of the second-order moment condition when Y is a general diffusion.

Remark III.2.4. When $\mu \in \mathcal{A}(T, y)$ is not centered around y (i.e. $\bar{\mu} \neq y$), the function H does not vanish in the constraint 2 of $\mathcal{A}(T, y)$. In this case, the measure μ can be embedded into Y under $\mathbb{P}^y$ by the following stopping rule τ: First wait until $\tau_{\bar{\mu}} = \inf\{t \in [0,\infty) \colon Y_t = \bar{\mu}\}$ and then embed μ in Y under $\mathbb{P}^{\bar{\mu}}$. To prove this, note that

$$q_z(x) = q_y(x) - q_y(z) - \frac{1}{2}(x - z) \left(\frac{\partial^+ q_y}{\partial x}(z) + \frac{\partial^- q_y}{\partial x}(z) \right), \quad x \in J, \tag{III.2.2}$$

where $\frac{\partial^+ q_y}{\partial x}$ and $\frac{\partial^- q_y}{\partial x}$ denote the right-hand side and left-hand side derivative of q_y, respectively. Let $a_n = -n$ if $\bar{\mu} > y$ (i.e. $l = -\infty$ by Remark III.2.1) and $a_n = n$ otherwise. Define $\tau_{\bar{\mu}, a_n} = \inf\{t \ge 0 \colon Y_t \notin (\bar{\mu} \wedge a_n, \bar{\mu} \vee a_n)\}$. Monotone convergence and Lemma 2.2 in [5] imply

$$\begin{aligned} \mathbb{E}^y[\tau_{\bar{\mu}}] &= \lim_{n\to\infty} \mathbb{E}^y[\tau_{\bar{\mu},a_n}] = \lim_{n\to\infty} \mathbb{E}^y\left[q_y\left(Y_{\tau_{\bar{\mu},a_n}}\right)\right] \\ &= q_y(\bar{\mu}) + \frac{1}{2} m(\{y\})|y - \bar{\mu}| + \mathbb{1}_{\{\bar{\mu}<y\}}(y - \bar{\mu}) m\big((y,\infty)\big) \\ &\quad + \mathbb{1}_{\{\bar{\mu}>y\}}(\bar{\mu} - y) m\big((-\infty, y)\big) \\ &= q_y(\bar{\mu}) + H(y, \bar{\mu}). \end{aligned} \tag{III.2.3}$$

(III.2.3) together with (III.2.2) yields that

$$\begin{aligned}\mathbb{E}^y[\tau] &= \mathbb{E}^y[\tau_{\bar{\mu}}] + \int_{\mathbb{R}} q_{\bar{\mu}}(x)\mu(dx) = q_y(\bar{\mu}) + H(y,\bar{\mu}) + \int_{\mathbb{R}} q_{\bar{\mu}}(x)\mu(dx)\\ &= \int_{\mathbb{R}} q_y(x)\mu(dx) + H(y,\bar{\mu}).\end{aligned}$$

Proposition III.2.2 allows to reformulate the stopping problem (III.1.3) as a linear problem on $\mathcal{M}$.

Corollary III.2.5. *We have*

$$V(T,y) = \sup_{\mu\in\mathcal{A}(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) \tag{III.2.4}$$

and for any optimal $\mu \in \mathcal{A}(T,y)$ there exists an optimal stopping time $\tau \in \mathcal{S}(T,y)$ in (III.1.3) *with $Y_\tau \sim \mu$ under $\mathbb{P}^y$. Conversely, let $\tau \in \mathcal{S}(T,y)$ be optimal in* (III.1.3) *and denote by μ the distribution of Y_τ under $\mathbb{P}^y$. Then $\mu \in \mathcal{A}(T,y)$ and μ is optimal in* (III.2.4).

Denote by $\mathcal{A}_3(T,y)$ the set of probability measures in $\mathcal{A}(T,y)$ which are a weighted sum of at most three Dirac measures. In the next three sections we show that the supremum in (III.2.4) is attained in $\mathcal{A}_3(T,y)$.

Theorem III.2.6. *We have*

$$V(T,y) = \sup_{\mu\in\mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx). \tag{III.2.5}$$

For the proof of Theorem III.2.6 we now introduce the balayage method developed by Chacon and Walsh in [17] for the Brownian motion.

III.3 Embedding Distributions with the Balayage Method

The aim of this section is to solve the Skorokhod embedding problem (SEP) for general diffusions with the balayage method introduced in [17] for the Brownian motion. The embedding stopping time τ is given by a limit of consecutive exit times and we show that the expected value of τ can be characterized in terms of the function q_y. We first recall some facts on potential functions associated to distributions.

III.3.1 Introduction to Potential Theory

Here we list some properties of potential functions in one dimension. Most of the proofs can be found e.g. in [16] and in Section 2.2 of [47].

Definition III.3.1. Let $\mu \in \mathcal{M}^1(\mathbb{R})$. The potential $u_\mu\colon \mathbb{R} \to \mathbb{R}$ of μ is defined by

$$u_\mu(x) = -\int_{\mathbb{R}} |x-z|\,\mu(dz).$$

Lemma III.3.2. *Let $\mu,\nu \in \mathcal{M}^1(\mathbb{R})$ with $\int_{\mathbb{R}} x\,\mu(dx) = w$. Then the potential satisfies the following properties.*

1. *$u_\mu \le u_{\delta_w}$.*

2. *The potential function u_μ is concave and Lipschitz-continuous with parameter 1.*

3. *If $\int_{\mathbb{R}} z\,\nu(dz) = \int_{\mathbb{R}} z\,\mu(dz)$, then we have $\lim_{|x|\to\infty} |u_\mu(x) - u_\nu(x)| = 0$.*

4. *$\mu(\{x\}) = \frac{1}{2}(\partial_- u_\mu - \partial_+ u_\mu)(x)$, where ∂_+ and ∂_- denote the right-hand and left-hand side derivative, respectively.*

5. *Let $a, b \in \mathbb{R}$ with $a < b$. u_μ is linear on $[a, b]$ if and only if $\mu\big((a, b)\big) = 0$.*

6. *If $u_\mu \le u_\nu$, then $\int_{\mathbb{R}} z\,\mu(dz) = \int_{\mathbb{R}} z\,\nu(dz)$, i.e. μ and ν have the same mean.*

7. *If $u_\mu = u_\nu$, then $\mu = \nu$.*

8. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{M}^1(\mathbb{R})$ such that $u_{\mu_n} \ge u_\nu$, $n \in \mathbb{N}$, and $(\mu_n)_{n\in\mathbb{N}}$ converges weakly to μ, i.e. $\mu_n \Rightarrow \mu$. Then for all $x \in \mathbb{R}$ it holds true that $\lim_{n\to\infty} u_{\mu_n}(x) = u_\mu(x)$.*

9. *Let $\mu_n \in \mathcal{M}^1(\mathbb{R})$, $n \in \mathbb{N}$, such that $\lim_{n\to\infty} u_{\mu_n}(x) = u_\mu(x)$ for all $x \in \mathbb{R}$ and $u_{\mu_n} \ge u_\nu$. Then $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.*

10. *Let $\mu_n \in \mathcal{M}^1(\mathbb{R})$, $n \in \mathbb{N}$, such that $\lim_{n\to\infty} u_{\mu_n}(x) = u_\mu(x)$ for all $x \in \mathbb{R}$. Let X_n and X be random variables on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \mathbb{P})$ with distribution μ_n and μ under $\mathbb{P}$, respectively. If $\lim_{n\to\infty} X_n = X$, $\mathbb{P}$-a.s., then $X_n \to X$ in $L^1(\mathbb{P})$ as $n \to \infty$.*

11. *Let $g\colon \mathbb{R} \to (-\infty, \infty]$ be a convex function. If $u_\mu \ge u_\nu$, then $\int_{\mathbb{R}} g(z)\mu(dz) \le \int_{\mathbb{R}} g(z)\nu(dz)$.*

Proof. We only prove Properties 4, 10 and 11.

4. For $\mu \in \mathcal{M}^1(\mathbb{R})$ with $\int_{\mathbb{R}} x\,\mu(dx) = w$ it holds that

$$\begin{aligned} u_\mu(x) &= -\int_{(-\infty,x)} (x - z)\mu(dz) - \int_{[x,\infty)} (z - x)\mu(dz) \\ &= x\left(2\mu\big([x,\infty)\big) - 1\right) + w - 2\int_{[x,\infty)} z\,\mu(dz). \end{aligned}$$

In particular, we conclude that

$$\partial_+ u_\mu(x) = 2\mu\big((x,\infty)\big) - 1, \qquad \partial_- u_\mu(x) = 2\mu\big([x,\infty)\big) - 1$$

and hence,

$$\mu(\{x\}) = \frac{1}{2}\big(\partial_- u_\mu(x) - \partial_+ u_\mu(x)\big).$$

10. Let μ_n, $\mu \in \mathcal{M}^1(\mathbb{R})$, $n \in \mathbb{N}$, such that $u_{\mu_n}(x) \xrightarrow[n\to\infty]{} u_\mu(x)$ for every $x \in \mathbb{R}$. Let X_n and X be random variables on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \mathbb{P})$ with distribution μ_n and μ under $\mathbb{P}$, respectively. Then

$$\mathbb{E}[|X_n|] = \int_{\mathbb{R}} |z|\,\mu_n\,(dz) = -u_{\mu_n}(0) \xrightarrow[n\to\infty]{} -u_\mu(0) = \mathbb{E}[|X|]. \tag{III.3.1}$$

Moreover, assume that $X_n \to X$, $\mathbb{P}$-a.s., as $n \to \infty$. Define $Y_n := |X| + |X_n| - |X_n - X|$. Observe that Y_n is non-negative and that $Y_n \to 2|X|$, $\mathbb{P}$-a.s., as $n \to \infty$. Fatou's lemma and (III.3.1) now imply that

$$\mathbb{E}[2|X|] = \mathbb{E}[\liminf_{n\to\infty} Y_n] \le \liminf_{n\to\infty} \mathbb{E}[Y_n] = 2\mathbb{E}[|X|] - \limsup_{n\to\infty} \mathbb{E}[|X_n - X|].$$

Hence, $\limsup_{n\to\infty} \mathbb{E}[|X_n - X|] \le 0 \le \liminf_{n\to\infty} \mathbb{E}[|X_n - X|]$. Therefore, we conclude that $X_n \xrightarrow[n\to\infty]{} X$ in $L^1(\mathbb{P})$.

Observe that (III.3.1) and Scheffé's Lemma (see e.g. [70], Lemma 5.10(ii)) also imply the statement.

11. Let $g\colon \mathbb{R} \to \mathbb{R}$ be a continuous, convex and piecewise linear function with finitely many kinks. More precisely, let

$$g(x) = \sum_{n=0}^{N} \mathbb{1}_{\{(x_n, x_{n+1}]\}}(x) g_n(x), \tag{III.3.2}$$

where $N \in \mathbb{N}$, $-\infty = x_0 < x_1 < ... < x_N < x_{N+1} = \infty$ and $g_n\colon \mathbb{R} \to \mathbb{R}$ are linear functions with slope $s_n \in \mathbb{R}$, $0 \le n \le N$, where $s_{n-1} < s_n$ and $g_{n-1}(x_n) = g_n(x_n)$ for $1 \le n \le N$. Therefore,

$$g(x) = g_0(x_1) + s_0(x - x_1) + \sum_{n=1}^{N} \mathbb{1}_{\{x > x_n\}}(s_n - s_{n-1})(x - x_n).$$

For $\mu, \nu \in \mathcal{M}^1(\mathbb{R})$ with $u_\nu \le u_\mu$ we have,

$$\begin{aligned}
&\int_{\mathbb{R}} g(z)\,\nu(dz) - \int_{\mathbb{R}} g(z)\,\mu(dz) \\
&\quad = \int_{\mathbb{R}} \big(g_0(x_1) + s_0(z - x_1)\big)\,\nu(dz) - \int_{\mathbb{R}} \big(g_0(x_1) + s_0(z - x_1)\big)\,\mu(dz) \\
&\qquad + \sum_{n=1}^{N} (s_n - s_{n-1}) \left(\int_{(x_n,\infty)} (z - x_n)\,\nu(dz) - \int_{(x_n,\infty)} (z - x_n)\,\mu(dz) \right) \\
&\quad = \sum_{n=1}^{N} (s_n - s_{n-1}) \left(\int_{(x_n,\infty)} (z - x_n)\,\nu(dz) - \int_{(x_n,\infty)} (z - x_n)\,\mu(dz) \right),
\end{aligned}$$

where we use that $\int_{\mathbb{R}} z\,\mu(dz) = \int_{\mathbb{R}} z\,\nu(dz)$ by Property 6 of Lemma III.3.2 because $u_\nu \le u_\mu$. Observe that

$$u_\mu(x) = -\int_{\mathbb{R}} |x - z|\,\mu(dz) = \int_{\mathbb{R}} z\,\mu(dz) - x - 2\int_{(x,\infty)} (z - x)\mu(dz).$$

Hence,

$$\int_{\mathbb{R}} g(z)\,\nu(dz) - \int_{\mathbb{R}} g(z)\,\mu(dz) = \frac{1}{2} \sum_{n=1}^{N} (s_n - s_{n-1}) \left(u_\mu(x_n) - u_\nu(x_n)\right) \ge 0,$$

because $s_n > s_{n-1}$ for $1 \le n \le N$ and $u_\mu \ge u_\nu$.

Now assume that g is an arbitrary convex function. Then there exists a sequence $(\tilde{g}_n)_{n \in \mathbb{N}}$ of functions of the form (III.3.2) such that $\tilde{g}_n \le \tilde{g}_{n+1}$ and $\tilde{g}_n \to g$ as $n \to \infty$. Monotone convergence yields

$$\int_{\mathbb{R}} g(z)\,\mu(dz) = \lim_{n\to\infty} \int_{\mathbb{R}} \tilde{g}_n(z)\,\mu(dz) \le \lim_{n\to\infty} \int_{\mathbb{R}} \tilde{g}_n(z)\,\nu(dz) = \int_{\mathbb{R}} g(z)\,\nu(dz).$$

□

The concept of the balayage of a probability measure μ on an interval (a, b), $a < b$, is essential for constructing a solution of the SEP for Brownian motion introduced by Chacon and Walsh [17]. The balayage of μ on (a, b) coincides with μ on $\mathbb{R} \backslash [a, b]$ and the mass of μ inside $[a, b]$ is shifted to the points a and b in such a way that the mean is preserved.

Definition III.3.3. Let $\mu \in \mathcal{M}^1(\mathbb{R})$ and $a < b$. The balayage $\mu^{(a,b)}$ of μ on the interval (a, b) is a probability measure on $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ defined by

$$\mu^{(a,b)}(A) = \begin{cases} \mu\big(A \cap \big((-\infty, a) \cup (b, \infty)\big)\big), & \text{if } a, b \notin A, \\ \mu\big(A \cap \big((-\infty, a) \cup (b, \infty)\big)\big) + \int_{[a,b]} \frac{b-x}{b-a}\,\mu(dx), & \text{if } a \in A, b \notin A, \\ \mu\big(A \cap \big((-\infty, a) \cup (b, \infty)\big)\big) + \int_{[a,b]} \frac{x-a}{b-a}\,\mu(dx), & \text{if } a \notin A, b \in A, \\ \mu\big(A \cup [a, b]\big), & \text{if } a, b \in A, \end{cases}$$

where $A \in \mathcal{B}(\mathbb{R})$.

Remark III.3.4. The balayage $\mu^{(a,b)}$ is well-defined. We conclude from the definition that

$$\mu^{(a,b)}(\{a\}) = \int_{[a,b]} \frac{b-x}{b-a}\,\mu(dx), \qquad \mu^{(a,b)}(\{b\}) = \int_{[a,b]} \frac{x-a}{b-a}\,\mu(dx), \tag{III.3.3}$$

$$\mu^{(a,b)}\big|_{(-\infty,a)\cup(b,\infty)} = \mu\big|_{(-\infty,a)\cup(b,\infty)}, \qquad \mu^{(a,b)}(A) = 0, \quad \text{if } A \in \mathcal{B}(\mathbb{R}),\ A \subset (a,b).$$

Lemma III.3.5. *Let $\mu \in \mathcal{M}^1(\mathbb{R})$ and $a < b$. The potential $u_{\mu^{(a,b)}}$ of the balayage of μ on (a,b) satisfies*

1. *$u_{\mu^{(a,b)}}(x) = u_\mu(x)$ for all $x \in \mathbb{R}\backslash(a,b)$,*
2. *$u_{\mu^{(a,b)}}$ is linear on $[a,b]$,*
3. *$u_{\mu^{(a,b)}} \le u_\mu$.*

Proof. 1. Let $x \le a$. Combining $\mu^{(a,b)} = \mu$ on $\mathbb{R}\backslash[a,b]$, $\mu^{(a,b)}\big((a,b)\big) = 0$ and (III.3.3) entails that

$$\begin{aligned} u_\mu(x) - u_{\mu^{(a,b)}}(x) &= \int_{[a,b]} (z-x)\,\mu^{(a,b)}(dz) - \int_{[a,b]} (z-x)\,\mu(dz) \\ &= \int_{[a,b]} \left(\frac{(a-x)(b-z)}{b-a} + \frac{(b-x)(z-a)}{b-a} - (z-x) \right) \mu(dz) = 0. \end{aligned}$$

Similarly, one can show that $u_{\mu^{(a,b)}}(x) = u_\mu(x)$ for all $x \ge b$.

2. Since $\mu^{(a,b)}\big((a,b)\big) = 0$, Property 5 in Lemma III.3.2 implies that $u_{\mu^{(a,b)}}$ is linear on $[a,b]$.

3. Using that u_μ is concave and continuous by Property 2, Lemma III.3.2, we conclude from 1. and 2. that $u_{\mu^{(a,b)}} \le u_\mu$.

□

III.3.2 The Embedding Method by Chacon and Walsh

Let $\mu \in \mathcal{M}^1(J) \subseteq \mathcal{M}^1(\mathbb{R})$ be a probability measure with $\int_\mathbb{R} q_y(x)\mu(dx) < \infty$. In particular, we have $\int_\mathbb{R} q_{\bar\mu}(x)\mu(dx) < \infty$ by (III.2.2). Using the ideas of Chacon and Walsh in [17] we construct a stopping time τ such that $Y_\tau \sim \mu$ under $\mathbb{P}^{\bar\mu}$ and $\mathbb{E}^{\bar\mu}[\tau] = \int_\mathbb{R} q_{\bar\mu}(x)\mu(dx)$. In addition, $\mathbb{E}^{\bar\mu}[\tau] \le \mathbb{E}^{\bar\mu}[\rho]$ for all embeddings ρ of μ in Y under $\mathbb{P}^{\bar\mu}$. In the following we construct an increasing sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ and measures $(\mu_n)_{n\in\mathbb{N}}$ such that

- $u_{\mu_n} \ge u_{\mu_{n+1}} \ge u_\mu$ for all $n \in \mathbb{N}$,
- $\lim_{n\to\infty} u_{\mu_n}(x) = u_\mu(x)$ for all $x \in \mathbb{R}$,
- $Y_{\tau_n} \sim \mu_n$ under $\mathbb{P}^{\bar\mu}$

and let $\tau = \lim_{n\to\infty} \tau_n$. Then the results for the potentials show that $Y_\tau \sim \mu$ under $\mathbb{P}^{\bar\mu}$. Proceed as follows:

Algorithm III.3.6. Let A be a countable dense subset of $\mathbb{R}$.

1. Let $\mu_0 = \delta_{\bar\mu}$, $\tau_0 = 0$ and $j = 0$.

2. Let $\ell = 0$.

3. Let $I_\ell^j = \left[-2^j + \frac{\ell}{2^j}, -2^j + \frac{\ell+1}{2^j}\right)$ and set $n = \frac{1}{3}\left(2^{2j+1} + 1\right) + \ell$.

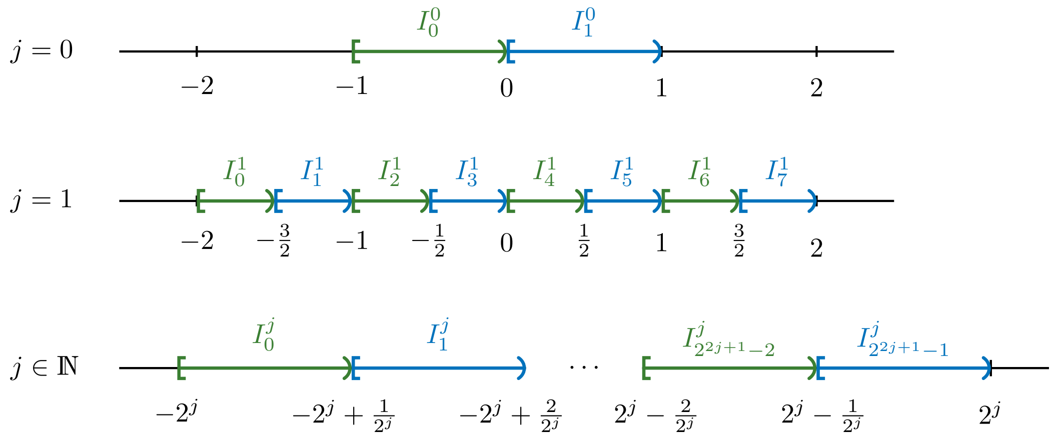


Figure III.2: The intervals I_ℓ^j for $j = 0$, $j = 1$ and $\ell \in \{0, ..., 2^{2j+1} - 1\}$ as well as the general case $j \in \mathbb{N}$.

a) If $u_{\mu_{n-1}}(x) > u_\mu(x)$ for some $x \in I_\ell^j$, choose a point $x_n \in A \cap I_\ell^j$ with $u_{\mu_{n-1}}(x_n) > u_\mu(x_n)$ and draw a tangent t_n to u_μ in x_n with slope $s_n \in \{\partial_+ u_\mu(x_n), \partial_- u_\mu(x_n)\}$. You can alternatively choose $x_n \in A \cap I_\ell^j$ such that $u_{\mu_{n-1}}(x_n) = u_\mu(x_n)$ but either $\partial_+ u_\mu(x_n) \neq \partial_+ u_{\mu_{n-1}}(x_n)$ or $\partial_- u_\mu(x_n) \neq \partial_- u_{\mu_{n-1}}(x_n)$. In this case draw the tangent t_n with slope $\partial_+ u_\mu(x_n)$ or $\partial_- u_\mu(x_n)$, respectively.

   Denote by a_n and b_n the intersection points of t_n and $u_{\mu_{n-1}}$. Let $\mu_n = \mu_{n-1}^{(a_n, b_n)}$ be the balayage of μ_{n-1} on (a_n, b_n) and $\tau_n = \tau_{n-1} + \rho(a_n, b_n) \circ \theta_{\tau_{n-1}}$, where $\rho(a, b) = \inf\{t \in [0, \infty) \colon Y_t \notin (a, b)\}$.

b) If $u_\mu = u_{\mu_{n-1}}$ on I_ℓ^j, set $\mu_n = \mu_{n-1}$ and $\tau_n = \tau_{n-1}$.

4. If $\ell < 2^{2j+1} - 1$, set $\ell = \ell + 1$ and continue with step 3, else continue with step 5.

5. If there exists $x \in A$ with $u_{\mu_n}(x) > u_\mu(x)$, let $j = j + 1$ and continue with step 2; otherwise we have $u_{\mu_n} = u_\mu$ and τ_n embeds μ into Y. Set $\tau_k = \tau_n$ and $\mu_k = \mu$ for all $k \geq n$.

Figure III.2 depicts the intervals I_ℓ^j for $j = 0, j = 1$ and sketches I_ℓ^j for $j \in \mathbb{N}$. Furthermore, the first steps of the construction described in Algorithm III.3.6 are illustrated in Figure III.3.

Remark III.3.7.

a) Observe that the slope s_n of the tangent t_n in x_n to u_μ in Algorithm III.3.6 satisfies $s_n \in (-1, 1)$. Indeed, since u_μ is concave, the left- and right-hand side derivatives $\partial_- u_\mu$ and $\partial_+ u_\mu$ are decreasing on $\mathbb{R}$ with $\partial_- u_\mu \geq \partial_+ u_\mu$ and their absolute value is bounded by the Lipschitz constant 1. Hence, $s_n \in [-1, 1]$. In the following we show $s_n < 1$. Assume, on the contrary, that $s_n = 1$ for some $n \in \mathbb{N}$. Then it follows that $\partial_- u_\mu(x_n) = 1$. Define

$$x_- = \sup\{x \leq \bar{\mu} : \partial_- u_\mu(x) = 1\}$$

with $\sup \emptyset := -\infty$. Hence $x_n \leq x_-$. Since $|u_\mu(x) + |x - \bar{\mu}|| \to 0$ as $x \to -\infty$ and $u_\mu \leq -|x - \bar{\mu}|$ by Properties 3 and 1 of Lemma III.3.2 and since u_μ is continuous, we deduce that $u_\mu(x) = -|x - \bar{\mu}|$ for $x \leq x_-$. Recall that x_n either satisfies

(i) $u_{\mu_{n-1}}(x_n) > u_\mu(x_n)$ or
(ii) $u_{\mu_{n-1}}(x_n) = u_\mu(x_n)$ and $\partial_+ u_{\mu_{n-1}}(x_n) \neq \partial_+ u_\mu(x_n)$ or
(iii) $u_{\mu_{n-1}}(x_n) = u_\mu(x_n)$ and $\partial_- u_{\mu_{n-1}}(x_n) \neq \partial_- u_\mu(x_n)$.

Due to the construction of $u_{\mu_{n-1}}$ we have $u_{\mu_{n-1}}(x) = -|x - \bar{\mu}| = u_\mu(x)$ for all $x \leq x_-$. Hence, only (ii) is possible with $x_n = x_-$ and $\partial_+ u_\mu(x_n) = s_n = 1 > \partial_+ u_{\mu_{n-1}}(x_n)$. But this contradicts the fact that $u_{\mu_{n-1}}$ is piecewise linear and $u_{\mu_{n-1}} \geq u_\mu$. Therefore, $s_n < 1$. Similarly, one concludes that $s_n > -1$.

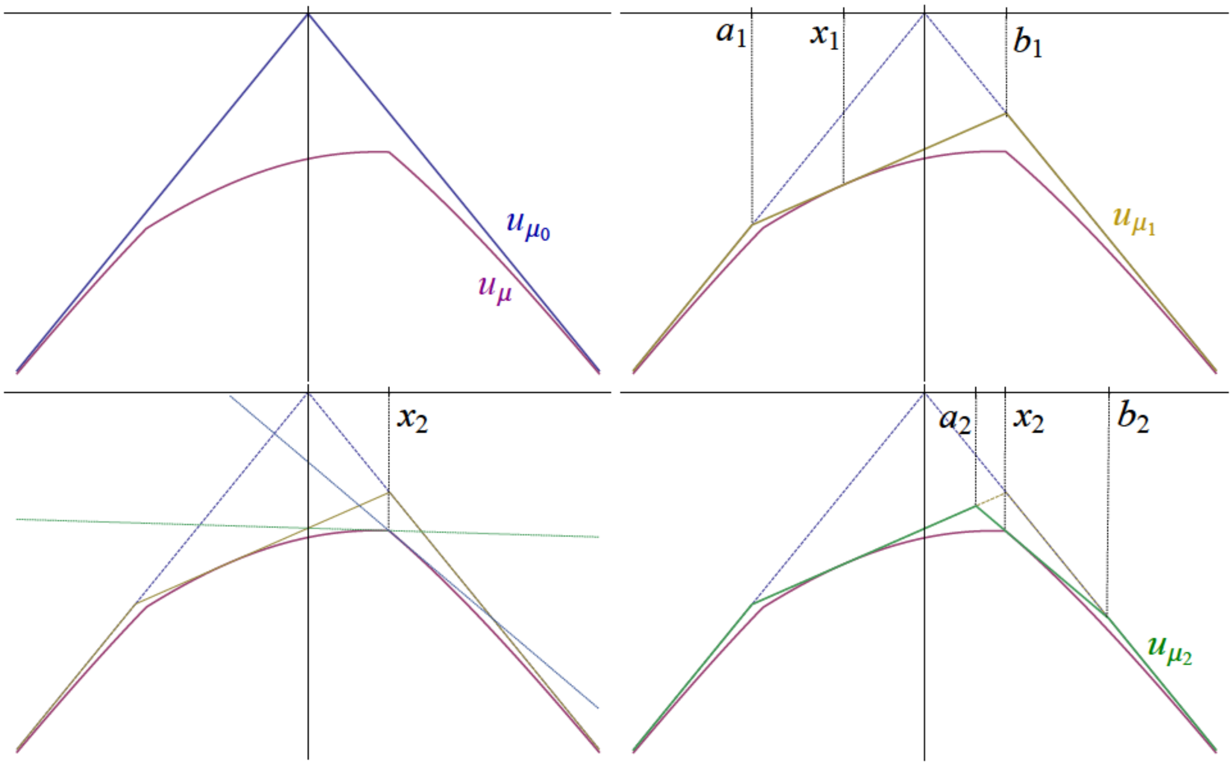


Figure III.3: The first two steps in the construction of a solution of the SEP according to Chacon and Walsh for a measure μ with kinks. In the second step (second line) we choose the right-hand side tangent.

b) As a consequence $u_{\mu_0} = u_{\delta_{\bar{\mu}}}$ and t_n have exactly two intersection points. Since $u_{\mu_{n-1}} \leq u_{\mu_0}$ and there exists a point $z \in \mathbb{R}$ with $t_n(z) < u_{\mu_{n-1}}(z)$, it follows that $u_{\mu_{n-1}}$ and t_n intersect in at least two points. The concave function $u_{\mu_{n-1}}$ and the affine function t_n have at most two points of intersection. Thus, the intersection points a_n and b_n obtained in Algorithm III.3.6 exist and a_n differs from b_n.

c) The intersection points a_n and b_n of the tangent t_n and the potential $u_{\mu_{n-1}}$ are contained in the state space J. Indeed, μ has support in J, thus, $u_\mu = u_{\delta_{\bar{\mu}}}$ on $\mathbb{R}\backslash J$; in particular $u_{\mu_{n-1}} = u_\mu$ on $\mathbb{R}\backslash J$. If we draw a tangent in $x_n \in J$ to u_μ in Algorithm III.3.6, then the slope of the tangent takes a value in $(-1, 1)$ by a). The concavity of u_μ implies that $t_n \geq u_\mu$ and thus, $t_n \geq u_{\mu_{n-1}}$ on $\mathbb{R}\backslash J$. Furthermore, the left- and right-hand side derivatives of $u_{\mu_{n-1}}$ decrease on J, with $\partial_- u_{\mu_{n-1}}(l) = -1$ if $l > -\infty$ and $\partial_+ u_{\mu_{n-1}}(r) = 1$ if $r < \infty$. Hence, $t_n > u_{\mu_{n-1}}$ on $\mathbb{R}\backslash J$ and $a_n, b_n \in J$.

Now we prove that the sequence $(u_n)_{n\in\mathbb{N}}$ constructed in Algorithm III.3.6 converges to u_μ as $n \to \infty$.

Lemma III.3.8. *Let $\mu \in \mathcal{M}^1$ with $\int_{\mathbb{R}} q_{\bar{\mu}}(x)\mu(dx) < \infty$. Let $(u_{\mu_n})_{n\in\mathbb{N}}$ be the potentials constructed in Algorithm III.3.6. Then it holds that $\lim_{n\to\infty} u_{\mu_n}(x) = u_\mu(x)$ uniformly in $x \in \mathbb{R}$.*

Proof. Let $\varepsilon > 0$. Since $u_\mu \leq u_{\delta_{\bar{\mu}}}$, Property 3 of Lemma III.3.2 implies that there exists $C \in [0, \infty)$ such that

$$\left|u_\mu(x) - u_{\delta_{\bar{\mu}}}(x)\right| < \varepsilon \qquad \text{for all } x \in \mathbb{R} \text{ with } |x| > C.$$

Furthermore, we have $u_\mu \leq u_{\mu_n} \leq u_{\delta_{\bar{\mu}}}$, $n \in \mathbb{N}$, by construction. Therefore, for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$ with $|x| > C$

$$|u_{\mu_n}(x) - u_\mu(x)| < \varepsilon. \tag{III.3.4}$$

Now we derive an upper bound for the difference between u_μ and u_{μ_n} on $[-C,C]$ for n sufficiently large. Let $x,z \in \mathbb{R}$ with $x<z$ and denote by t_x and t_z the (right- or left-hand side) tangents to u_μ in x and z, respectively. Property 2 of Lemma III.3.2 implies that $\partial_+ u_\mu(w), \partial_- u_\mu(w) \in [-1,1]$ for $w \in \mathbb{R}$. Hence, it follows that for all $w \in [x,z]$

$$0 \leq \big(t_x(w) \wedge t_z(w)\big) - u_\mu(w) \leq 2(z-x). \tag{III.3.5}$$

Let $j \in \mathbb{N}$ with $2^j > \max\left\{2(C+1), \frac{8}{\varepsilon}\right\}$ and let $m \geq \frac{1}{3}\left(2^{2j+1}+1\right)$. To simplify notation, let $k = \frac{1}{3}(2^{2j-1}+1)$ and for $0 \leq \ell \leq 3k-2 = 2^{2j-1}-1$ denote by $z_\ell = x_{k+\ell}$ the points where we draw the nth tangent, $n \in \left\{\frac{1}{3}\left(2^{2j-1}+1\right), \ldots, \frac{1}{3}\left(2^{2j+1}-2\right)\right\}$ in Algorithm III.3.6. Then

$$z_\ell \in \left[-2^{j-1} + \frac{\ell}{2^{j-1}}, -2^{j-1} + \frac{\ell+1}{2^{j-1}}\right)$$

and it follows from (III.3.5) that for $x \in [z_\ell, z_{\ell+1}]$, $0 \leq \ell \leq 3k-3$, we have

$$0 \leq u_{\mu_n}(x) - u_\mu(x) \leq \big(t_{z_\ell}(x) \wedge t_{z_{\ell+1}}(x)\big) - u_\mu(x) \leq 2(z_{\ell+1} - z_\ell) \leq 2 \cdot 2 \cdot 2^{-(j-1)} < \varepsilon \tag{III.3.6}$$

by the choice of j. Now observe that $[-C,C] \subseteq [z_0, z_{3k-2}]$. Therefore, (III.3.4) and (III.3.6) imply that for all $m \geq \frac{1}{3}\left(2^{2j+1}+1\right)$

$$\sup_{x\in\mathbb{R}} |u_{\mu_m}(x) - u_\mu(x)| < \varepsilon.$$

□

Now we show that the stopping time $\tau = \lim_{n\to\infty} \tau_n$ embeds μ into Y under $\mathbb{P}^y$ and compute the expectation of τ.

Proposition III.3.9. *Let $y \in (l,r)$. Let $\mu \in \mathcal{M}^1$ with $\bar{\mu} = y$ and $\int_{\mathbb{R}} q_y(x)\,\mu(dx) < \infty$. Let $(\mu_n)_{n\in\mathbb{N}}$ and $(\tau_n)_{n\in\mathbb{N}}$ be the probability measures and stopping times constructed in Algorithm III.3.6. Let $\tau = \lim_{n\to\infty}\tau_n$. Then*

1. *$Y_{\tau_n} \sim \mu_n$ under $\mathbb{P}^y$,*
2. *$\mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\,\mu(dx)$,*
3. *$Y_{\tau_n} \xrightarrow[n\to\infty]{} Y_\tau$, $\mathbb{P}^y$-a.s. and in $L^1(\mathbb{P}^y)$,*
4. *$Y_\tau \sim \mu$ under $\mathbb{P}^y$,*
5. *$q_y(Y_{\tau_n}) \xrightarrow[n\to\infty]{} q_y(Y_\tau)$, $\mathbb{P}^y$-a.s. and in $L^1(\mathbb{P}^y)$.*

Proof. 1. Let $n \in \mathbb{N}$. First observe that μ_n has at most $n+1$ mass point and all mass points are contained in J, see Remark III.3.7 c). Moreover, it holds that $\overline{\mu_n} = \bar{\mu}$, thus, $\mu_n \in \mathcal{M}^1$. In order to apply Lemma A.2.1 from Appendix A.2, we have to show that τ_{n-1} for $n \in \mathbb{N}$, $n \geq 1$, is integrable with respect to $\mathbb{P}^y$. This follows inductively from the strong Markov property and the fact that $\mathbb{E}^x[\rho(a,b)] < \infty$ for all $a,b,x \in J$, $a<x<b$ (cf. Proposition 3.1, Chapter VII in [56]). Using that $(Y_{t\wedge\tau_{n-1}})_{t\in[0,\infty)}$ is bounded for all $n \in \mathbb{N}$, Lemma A.2.1 from Appendix A.2 guarantees that Y_{τ_n} has distribution $\mu_n = \mu_{n-1}^{(a_n,b_n)}$ under $\mathbb{P}^y$.

2. Let $\vartheta_k = \inf\{t \in [0,\infty) \colon q_y(Y_t) \geq k\} \wedge k$. Then $(\vartheta_k)_{k\in\mathbb{N}}$ is a localizing sequence for the $\mathbb{P}^y$-local martingale $\big(q_y(Y_t) - \big(t \wedge \tau(l,r)\big)\big)_{t\in[0,\infty)}$. By construction we have

$$Y_{\tau_n \wedge t} \in \left[\min_{1\leq\ell\leq n} a_\ell, \max_{1\leq\ell\leq n} b_\ell\right] \subseteq J$$

for every $t \in [0,\infty)$ and thus, $q_y(Y_{\tau_n\wedge\vartheta_k})$ is bounded for every $n \in \mathbb{N}$.

The monotone convergence Theorem and Fatou's lemma imply that

$$\mathbb{E}^y[\tau] = \liminf_{n\to\infty} \lim_{k\to\infty} \mathbb{E}^y[\tau_n \wedge \vartheta_k] = \liminf_{n\to\infty} \lim_{k\to\infty} \mathbb{E}^y[q_y(Y_{\tau_n\wedge\vartheta_k})] = \liminf_{n\to\infty} \mathbb{E}^y\left[\lim_{k\to\infty} q_y(Y_{\tau_n\wedge\vartheta_k})\right]$$

$$= \liminf_{n\to\infty} \mathbb{E}^y[q_y(Y_{\tau_n})] \geq \mathbb{E}^y\left[\liminf_{n\to\infty} q_y(Y_{\tau_n})\right] = \mathbb{E}^y[q_y(Y_\tau)]$$

$$= \int_{\mathbb{R}} q_y(x)\mu(dx).$$

On the other hand, observe that q_y is convex for every $y \in (l, r)$. Using the previous calculation, $Y_{\tau_n} \sim \mu_n$ and Property 11 of Lemma III.3.2 we obtain that

$$\mathbb{E}^y[\tau] = \lim_{n\to\infty} \mathbb{E}^y[\tau_n] = \lim_{n\to\infty} \int_{\mathbb{R}} q_y(x)\mu_n(dx) \leq \int_{\mathbb{R}} q_y(x)\mu(dx).$$

Thus, we have shown that

$$\mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\mu(dx).$$

In particular, we conclude that $\tau < \infty$, $\mathbb{P}^y$-a.s.

3. Since $\tau_n \to \tau$ as $n \to \infty$, $\tau < \infty$, $\mathbb{P}^y$-a.s., and Y has continuous paths, we deduce that $Y_{\tau_n} \to Y_\tau$, $\mathbb{P}^y$-a.s. Moreover, Lemma III.3.8 and Property 10 in Lemma III.3.2 imply that $Y_{\tau_n} \to Y_\tau$ in $L^1(\mathbb{P}^y)$ as $n \to \infty$.

4. By *3.* the random variables Y_{τ_n} converge in distribution to Y_τ as $n \to \infty$. Thus, *1.* implies that $\mu_n \Rightarrow \nu$, where ν denotes the distribution of Y_τ under $\mathbb{P}^y$. On the other hand, Lemma III.3.8 and Property 9 of Lemma III.3.2 yield that $\mu_n \Rightarrow \mu$ as $n \to \infty$. So we deduce that $Y_\tau \sim \nu = \mu$ under $\mathbb{P}^y$.

5. First recall that q_y is continuous on J. Since μ_n and μ have support in J, it follows that Y_{τ_n}, $Y_\tau \in J$, $\mathbb{P}^y$-a.s., for all $n \in \mathbb{N}$. Moreover, $Y_{\tau_n} \to Y_\tau$, $\mathbb{P}^y$-a.s., by *3.*, hence, $\lim_{n\to\infty} q_y(Y_{\tau_n}) = q_y(Y_\tau)$, $\mathbb{P}^y$-a.s. For the $L^1(\mathbb{P}^y)$-convergence we use the same argument as in the proof of Property 10 in Lemma III.3.2. With *2.* applied to both $\mathbb{E}^y[\tau_n]$ and $\mathbb{E}^y[\tau]$ we conclude from monotone convergence that

$$\lim_{n\to\infty} \mathbb{E}^y[q_y(Y_{\tau_n})] = \lim_{n\to\infty} \int_{\mathbb{R}} q_y(x)\,\mu_n(dx) = \lim_{n\to\infty} \mathbb{E}^y[\tau_n] = \mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\,\mu(dx) = \mathbb{E}^y[q_y(Y_\tau)].$$

Hence,

$$\begin{aligned}\mathbb{E}^y[2q_y(Y_\tau)] &= \mathbb{E}^y\left[\liminf_{n\to\infty}\big(q_y(Y_\tau) + q_y(Y_{\tau_n}) - |q_y(Y_\tau) - q_y(Y_{\tau_n})|\big)\right]\\ &\leq \liminf_{n\to\infty} \mathbb{E}^y\left[q_y(Y_\tau) + q_y(Y_{\tau_n}) - |q_y(Y_\tau) - q_y(Y_{\tau_n})|\right]\\ &= \mathbb{E}^y[2q_y(Y_\tau)] - \limsup_{n\to\infty} \mathbb{E}^y[|q_y(Y_\tau) - q_y(Y_{\tau_n})|].\end{aligned}$$

Therefore, $q_y(Y_{\tau_n}) \xrightarrow[n\to\infty]{} q_y(Y_\tau)$ in $L^1(\mathbb{P}^y)$. □

Remark III.3.10. The stopping time τ constructed in Algorithm III.3.6 is minimal in the sense that $\tau = \rho$, $\mathbb{P}^{\bar{\mu}}$-a.s, for every stopping time ρ with $\rho \leq \tau$, $\mathbb{P}^{\bar{\mu}}$-a.s., that embeds μ in Y under $\mathbb{P}^{\bar{\mu}}$, see Theorem 2.4 in [32]. Moreover, all embedding stopping times ρ for μ satisfy $\mathbb{E}^{\bar{\mu}}[\rho] \geq \int_{\mathbb{R}} q_{\bar{\mu}}(x)\mu(dx) = \mathbb{E}^{\bar{\mu}}[\tau]$. Hence, τ is an embedding for μ with minimal expectation.

Now we examine the case when $\mu \in \bigcup_{T\in[0,\infty)} \mathcal{A}(T, y)$ and $\bar{\mu} \neq y$. By [32] a measure $\mu \in \mathcal{M}^1$ with $\bar{\mu} \neq y$ can be embedded in Y under $\mathbb{P}^y$ with an integrable stopping time τ if and only if $\mathbb{E}^y[\tau_{\bar{\mu}}]$ and $\int_{\mathbb{R}} q_{\bar{\mu}}(x)\mu(dx)$ are both finite, cf. Remark III.2.4. Recall that $\tau_{\bar{\mu}}$ denotes the first time when the process Y hits the point $\bar{\mu}$. The idea is to wait until the process Y hits the mean $\bar{\mu}$ for the first time and then to construct a minimal stopping time ϑ embedding μ in Y under $\mathbb{P}^{\bar{\mu}}$ using the balayage technique described in Algorithm III.3.6. We define

$$\tau = \tau_{\bar{\mu}} + \vartheta \circ \theta_{\tau_{\bar{\mu}}}.$$

Then due to the strong Markov property of Y we obtain that $Y_\tau \sim \mu$ under $\mathbb{P}^y$. In addition, Remark III.2.4 implies that

$$\mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\mu(dx) + H(y, \bar{\mu}).$$

Theorem 2.4 in [32] implies that τ has minimal expectation among all stopping times embedding the distribution μ in Y under $\mathbb{P}^y$. Summarizing, we have proven

Proposition III.3.11. *Let $\mu \in \bigcup_{T\in[0,\infty)} \mathcal{A}(T, y)$. Let $(\mu_n)_{n\in\mathbb{N}}$ and $(\vartheta_n)_{n\in\mathbb{N}}$ be the probability measures and stopping times constructed in Algorithm III.3.6. Let $\vartheta = \lim_{n\to\infty} \vartheta_n$ and define $\tau_n = \tau_{\bar{\mu}} + \vartheta_n \circ \theta_{\tau_{\bar{\mu}}}$ and $\tau = \tau_{\bar{\mu}} + \vartheta \circ \theta_{\tau_{\bar{\mu}}}$. Then*

1. *$\tau_n \xrightarrow[n\to\infty]{} \tau$, $\mathbb{P}^y$-a.s and $\tau < \infty$, $\mathbb{P}^y$-a.s,*
2. *$Y_{\tau_n} \sim \mu_n$, $Y_\tau \sim \mu$ under $\mathbb{P}^y$,*
3. *$\mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\, \mu(dx) + H(y, \bar{\mu})$,*
4. *$Y_{\tau_n} \xrightarrow[n\to\infty]{} Y_\tau$ and $q_y(Y_{\tau_n}) \xrightarrow[n\to\infty]{} q_y(Y_\tau)$, $\mathbb{P}^y$-a.s. and in $L^1(\mathbb{P}^y)$.*

III.4 Reduction to Atomic Measures

In this section we prove two main results of this chapter, formulated in Theorem III.1.5 and Theorem III.2.6. According to Corollary III.2.5 the optimal stopping problem (III.1.3) is equivalent to a measure optimization. In a first step we show by using the balayage method that for the measure optimization it is enough to consider only probability measures that are weighted sums of finitely many Dirac measures. In a second step we prove that it is sufficient to consider only measures in $\mathcal{A}_3(T, y)$, i.e. measures with at most 3 mass points. The claim of Theorem III.1.5 follows then by employing the embedding method of Section III.3.2.

In this section we impose the following stronger assumptions on f.

Assumption (B1). For every $y \in (l, r)$ there exists $C(y) \in [0, \infty)$ such that

$$|f(x)| \le C(y)\big(1 + q_y(x)\big), \qquad x \in J.$$

Assumption (B2). The set of discontinuity points $\mathcal{D} = \{x \in J \colon f \text{ is not continuous in } x\}$ is at most countable.

Recall that Assumption **(B1)** guarantees that the value function is finite, see Remark III.1.4. In Section III.5 we extend Theorem III.1.5 and Theorem III.2.6 to Borel-measurable payoff functions f that satisfy the weaker Assumption (**B**).

Remark III.4.1. If the process Y is a solution to the SDE $dY_t = \sigma(Y_t)dW_t$, $Y_0 = y$, where $\sigma\colon J \to \mathbb{R}$ is Lipschitz-continuous and W is a Brownian motion, then $q_y(x) = \int_y^x \int_y^z \frac{2}{\sigma^2(w)} dw\, dz$ (cf. Remark III.1.2). Thus, $q_y(x) = \tilde{q}_y(x - y)$, where $\tilde{q}_y$ denotes the function defined in (II.3.1). In particular, the growth condition in Assumption **(B1)** and in Proposition II.3.1 coincide.

III.4.1 Reduction to finitely many Atoms

Denote by $\mathcal{A}_a(T,y)$ the set of all probability measures in $\mathcal{A}(T,y)$ that can be written as a weighted sum of finitely many Dirac measures. We now prove that in the optimization problem (III.2.4) the supremum is attained in the set $\mathcal{A}_a(T,y)$.

Lemma III.4.2. *We have*

$$\sup_{\mu\in\mathcal{A}(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu\in\mathcal{A}_a(T,y)} \int_{\mathbb{R}} f(x)\mu(dx). \tag{III.4.1}$$

Proof. Since $\mathcal{A}_a(T,y) \subset \mathcal{A}(T,y)$, the left-hand side of (III.4.1) is greater than or equal to the right-hand side. In order to prove the reverse inequality let $\mu \in \mathcal{A}(T,y)$. Observe that $\mu \in \mathcal{A}(T,y)$ if and only if $\mu \in \mathcal{A}\big(T - H(y,\bar{\mu}) - q_y(\bar{\mu}), \bar{\mu}\big)$, cf. Remark III.2.4. Here we set $\mathcal{A}(S,x) = \emptyset$ if $S < 0$ or $x \notin J$ and $\mathcal{A}(S,x) = \{\delta_x\}$, if $x \in J\backslash(l,r)$ and $S \in [0,\infty)$. Therefore, we can assume that $\bar{\mu} = y$.

It is enough to show that there exists a sequence of probability measures $\mu_n \in \mathcal{A}_a(T,y)$ such that

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\mu_n(dx) = \int_{\mathbb{R}} f(x)\mu(dx). \tag{III.4.2}$$

First suppose that the set $\mathcal{D}$ of discontinuity points of f is infinite and denote by $d_1, d_2, \ldots$ the elements of $\mathcal{D}$. According to Algorithm III.3.6 with $A = \mathcal{D} \cup \mathbb{Q}$ we construct a sequence of potentials u_n, $n \in \mathbb{N}$, with

(1) $u_n \geq u_{n+1} \geq u_\mu$,

(2) $\lim_{n\to\infty} u_n(x) = u_\mu(x)$ for all $x \in \mathbb{R}$.

Then choose a subsequence, also denoted by $(u_n)_{n\in\mathbb{N}}$, such that in addition the following properties are satisfied:

(3) $u_n(d_i) = u_\mu(d_i)$ for all $i \in \{1,\ldots,n\}$,

(4) $\partial_+ u_n(d_i) = \partial_+ u_\mu(d_i)$ for all $i \in \{1,\ldots,n\}$,

(5) $\partial_- u_n(d_i) = \partial_- u_\mu(d_i)$ for all $i \in \{1,\ldots,n\}$.

We denote by μ_n the unique probability measure associated to u_n, $n \in \mathbb{N}$. Observe that each μ_n belongs to $\mathcal{A}_a(T,y)$, because μ_n is the kth balayage of the Dirac measure δ_y for some $k \in \mathbb{N}$ and it holds that $\int_{\mathbb{R}} q_y(x)\mu_n(dx) \leq \int_{\mathbb{R}} q_y(x)\mu(dx)$, $n \in \mathbb{N}$, by Property 11 of Lemma III.3.2. Moreover, it follows from Property 4 in Lemma III.3.2 that for $i \leq n$ we have

$$\mu_n(\{d_i\}) = \mu(\{d_i\}).$$

Let $(\tau_n)_{n\in\mathbb{N}}$ be the sequence of increasing stopping times embedding the distribution μ_n into Y under $\mathbb{P}^y$, which are constructed in Algorithm III.3.6. We write $X_n = Y_{\tau_n}$ and $X = Y_\tau$, where $\tau = \lim_{n\to\infty} \tau_n$. $(X_n)_{n\in\mathbb{N}}$ and $\big(q_y(X_n)\big)_{n\in\mathbb{N}}$ converge to X and $q_y(X)$, respectively, $\mathbb{P}^y$-almost surely and in $L^1(\mathbb{P}^y)$ by Lemma III.3.9. Consequently, the sequence $\big(q(X_n)\big)_{n\in\mathbb{N}}$ is uniformly integrable. The growth condition for f in Assumption **(B1)** entails that $\big(f(X_n)\big)_{n\in\mathbb{N}}$ is also uniformly integrable.

Notice that on $\{X \notin \mathcal{D}\}$ the sequence $\big(f(X_n)\big)_{n\in\mathbb{N}}$ converges $\mathbb{P}^y$-a.s. to $f(X)$, since f is continuous on $\mathbb{R}\setminus\mathcal{D}$. Moreover, on the event $\{X \in \mathcal{D}\}$ the sequence $(X_n)_{n\in\mathbb{N}}$ is constant equal to X eventually, see Corollary A.2.3 in Appendix A.2. Therefore, also on $\{X \in \mathcal{D}\}$ we have $\mathbb{P}^y$-a.s. convergence of $\big(f(X_n)\big)_{n\in\mathbb{N}}$ to $f(X)$. To sum up, the sequence $\big(f(X_n)\big)_{n\in\mathbb{N}}$ converges $\mathbb{P}^y$-a.s. to $f(X)$ and together with uniform integrability this yields (III.4.2).

Finally consider the case where $\mathcal{D} = \{d_1,\ldots,d_N\}$ is finite. In this case one can choose the subsequence of potentials $(u_n)_{n\in\mathbb{N}}$ such that the Properties (1), (2) are satisfied and for all $n \geq N$ we have

(3') $u_n(d_i) = u_\mu(d_i)$ for all $i \in \{1, \dots, N\}$,

(4') $\partial_+ u_n(d_i) = \partial_+ u_\mu(d_i)$ for all $i \in \{1, \dots, N\}$,

(5') $\partial_- u_n(d_i) = \partial_- u_\mu(d_i)$ for all $i \in \{1, \dots, N\}$.

The rest of the proof is similar to the proof of the first case. □

III.4.2 Reduction to 3 Atoms

We use here the methodology developed by Hoeffding [33] to show that the value of the optimal stopping problem (III.2.4) is attained in $\mathcal{A}_3(T, y)$.

Proposition III.4.3. *We have*

$$\sup_{\mu \in \mathcal{A}_a(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu \in \mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

Proof. The proof follows the lines of the proof of Theorem 2.1 in [33]. Let $\mu \in \mathcal{A}_a(T, y)$ be an arbitrary measure with more than 3 atoms and let $\bar{\mu} = \int_{\mathbb{R}} x\, \mu(dx)$. In the following we exhibit a measure $\widetilde{\nu} \in \mathcal{A}_3(T, y)$ such that $\int_{\mathbb{R}} f(x)\widetilde{\nu}(dx) \geq \int_{\mathbb{R}} f(x)\mu(dx)$ and $\int_{\mathbb{R}} x\, \widetilde{\nu}(dx) = \bar{\mu}$. By assumption, μ takes the form

$$\mu = \sum_{j=1}^{k} p_j \delta_{a_j},$$

where $k \geq 4$, $a_1 < a_2 < \cdots < a_k$ are k real numbers, $p_j > 0, 1 \leq j \leq k$, $\sum_{j=1}^k p_j = 1$, $\sum_{j=1}^k p_j a_j = \bar{\mu}$ and $\sum_{j=1}^k p_j q_y(a_j) \leq T - H(y, \bar{\mu})$. For $c_j \in \mathbb{R}, 1 \leq j \leq k$, and $t \in \mathbb{R}$ we define

$$\nu = \sum_{j=1}^{k} (p_j + tc_j)\delta_{a_j}.$$

To ensure that $\nu \in \mathcal{A}(T, y)$, it is sufficient that

$$\begin{aligned} p_j + tc_j \in [0,1],\ 1 \leq j \leq k&, \\ c_1 + \cdots + c_k = 0&, \\ c_1 a_1 + \cdots + c_k a_k = 0&, \\ c_1 q_y(a_1) + \cdots + c_k q_y(a_k) = 0&. \end{aligned}$$

We want to find $t \in \mathbb{R}$ and $(c_1, \dots, c_k) \in \mathbb{R}^k$ satisfying the previous constraints and

$$\int_{\mathbb{R}} f(x)\nu(dx) - \int_{\mathbb{R}} f(x)\mu(dx) = t \sum_{j=1}^{k} c_j f(a_j) \geq 0.$$

Let

$$M = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ a_1 & a_2 & \cdots & a_k \\ q_y(a_1) & q_y(a_2) & \cdots & q_y(a_k) \\ f(a_1) & f(a_2) & \cdots & f(a_k) \end{pmatrix} \quad \text{and} \quad \lambda = \begin{cases} 1 & \text{if } rank(M) = 4, \\ 0 & \text{if } rank(M) < 4. \end{cases}$$

Then the equation

$$Mc = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \lambda \end{pmatrix}$$

has a non-trivial solution $(c_1, \dots, c_k)$ (not equal to the null vector). Since these values sum to 0, there is at least one j such that $c_j < 0$. Let $I_- = \big\{j \in \{1, \cdots, k\} \colon c_j < 0\big\}$ and $I_+ = \big\{j \in \{1, \cdots, k\} \colon c_j \geq 0\big\}$. We define

$$t = \min\left\{-\frac{p_j}{c_j} \colon j \in I_-\right\}.$$

The value t is well defined and positive. This ensures that at least one of the values $(p_j + tc_j)$ equals 0. Moreover, for $j \in I_-$ we have $p_j + tc_j \geq p_j - c_j \frac{p_j}{c_j} = 0$ and for $j \in I_+$ it holds that $p_j + tc_j \geq p_j \geq 0$. Therefore, $\{p_j + tc_j \colon 1 \leq j \leq k\}$ forms a sequence of non-negative reals whose sum is equal to one. This entails that $p_j + tc_j \in [0, 1]$ for all $1 \leq j \leq k$. Thus, we have constructed a probability measure $\nu \in \mathcal{A}_a(T, y)$ with at most $k - 1$ atoms and such that $\int_{\mathbb{R}} f(x)\nu(dx) \geq \int_{\mathbb{R}} f(x)\mu(dx)$. Repeating this procedure a finite number of times we construct a measure $\widetilde{\nu} \in \mathcal{A}_a(T, y)$ with at most 3 atoms and satisfying all our requirements. $\square$

With Theorem III.2.6 we can prove Theorem III.1.5.

Proof of Theorem III.1.5. Let $\mu \in \mathcal{A}_3(T, y)$ with exactly three mass points $a < c < d$. First assume that μ is centered around y.

Using the balayage method described in Section III.3, we construct consecutive exit times as follows:

- Let $x_1 = a$. Property 4, Lemma III.3.2, implies

$$t_1(x) := \partial_+ u_\mu(x_1)(x - x_1) + u_\mu(x_1) = \big(1 - 2\mu(\{a\})\big)(x - a) + a - \bar{\mu}.$$

 Then, with the notation of Algorithm III.3.6, we obtain

$$\begin{aligned} a_1 = a, \qquad b_1 &= \frac{\mu(\{c\})c + \mu(\{d\})d}{1 - \mu(\{a\})}, \\ u_{\mu_1}(x) &= -|x - \bar{\mu}| \wedge t_1(x), \\ \mu_1 &= \mu(\{a\})\delta_a + \big(1 - \mu(\{a\})\big)\delta_{b_1}, \\ \tau_1 &= \inf\{t \in [0, \infty) \colon Y_t \notin (a, b_1)\}. \end{aligned}$$

- Now choose $x_2 = d$ and

$$t_2(x) := \partial_- u_\mu(x_2)(x - x_2) + u_\mu(x_2) = \big(2\mu(\{d\}) - 1\big)(x - d) + \bar{\mu} - d.$$

 Here we obtain $a_2 = c, b_2 = d$, $u_{\mu_2} = u_\mu$, $\mu_2 = \mu$ and

$$\tau_2 = \tau_1 + \inf\{t \in [0, \infty) \colon Y_t \notin (c, d)\} \circ \theta_{\tau_1}.$$

Therefore, the stopping time τ_2 is an embedding of μ into Y under $\mathbb{P}^y$. Moreover, observe that $b_1 \in (c, d)$, because b_1 is a strict convex combination of c and d. Since $a < \bar{\mu}$, it holds that $b_1 = \big(\bar{\mu} - \mu(\{a\})a\big)/\big(1 - \mu(\{a\})\big) > y$. Hence, $b_1 \in \big(\max\{c, \bar{\mu}\}, d\big)$. Thus, τ_2 can be rewritten as

$$\tau_2 = \tau_{a,b}(0) + \mathbb{1}_{\left\{Y_{\tau_{a,b}(0)} = b\right\}} \inf\big\{t \in [0, \infty) \colon Y_{t + \tau_{a,b}(0)} \in \{c, d\}\big\}, \tag{III.4.3}$$

where $b = b_1$. Proposition III.3.9, 2, implies $\mathbb{E}^y[\tau_2] = \int_{\mathbb{R}} q_y(x)\mu(dx) \leq T$; hence $\tau_2 \in \mathcal{S}_3(T, y)$.

If μ has two mass points a and d, then $\tau = \inf\{t \in [0, \infty) \colon Y_t \notin (a, d)\} \in \mathcal{S}_3(T, y)$. The stopping time τ can be described as a consecutive exit time of the form (III.4.3). Indeed, set $b = c = d$ in (III.4.3). And similar, if $\mu = \delta_y$, then $\tau = \inf\{t \in [0, \infty) \colon Y_t = y\} \equiv 0 \in \mathcal{S}_3(T, y)$. Here we set $a = b = c = d = y$.

If $\mu \in \mathcal{A}_3(T, y)$ is not centered around y, then the first hitting time of μ is integrable with respect to $\mathbb{P}^y$ (see Theorem 2.4 in [32]). Thus, we wait until Y hits $\bar{\mu}$ for the first time and then continue as in the centered case, see Remark III.2.4. Assume that μ has exactly three mass points $a < c < d$, then the balayage method for the uncentered case yields that

$$\tau := \tau_2 \circ \theta_{\tau_{\bar{\mu}}} = \tau_{a,b}(\tau_{\bar{\mu}}) + \mathbb{1}_{\left\{Y_{\tau_{a,b}(\tau_{\bar{\mu}})}=b\right\}} \inf\left\{t \in [0,\infty) \colon Y_{t+\tau_{a,b}(\tau_{\bar{\mu}})} \in \{c,d\}\right\} \tag{III.4.4}$$

embeds μ into Y under $\mathbb{P}^y$ with $\mathbb{E}^y[\tau] = \int_{\mathbb{R}} q_y(x)\mu(dx) + H(y,\bar{\mu}) \leq T$ by Proposition III.3.11, 3. Similarly, if $\mu \in \mathcal{A}_3(T,y)$ has two mass points $a < d$, then for $b = c = d$ the stopping time τ in (III.4.4) belongs to $\mathcal{S}_3(T,y)$. Finally, let $\mu = \delta_a \in \mathcal{A}_3(T,y)$ with $a \neq y$. Then the stopping time constructed in Section III.3 is given by $\tau = \tau_{\bar{\mu}} = \tau_a$ which can be obtained by setting $b = c = d = a$ in (III.4.4). □

Remark III.4.4. Observe that a measure $\mu \in \mathcal{A}(T,y)$ with exactly three mass points can be embedded in Y under $\mathbb{P}^y$ with different consecutive exit times. Indeed, let

$$\mu = \mu(\{a\})\delta_a + \mu(\{c\})\delta_c + \mu(\{d\})\delta_d \in \mathcal{A}(T,y)$$

with $a < c < d \in J$ and $\mu(\{a\}), \mu(\{c\}), \mu(\{d\}) \in (0,1)$. Now choose $p \in \big(\mu(\{a\}), \mu(\{a\}) + \mu(\{c\})\big)$ and set

$$b_1 = c - \frac{(c-a)\mu(\{a\})}{p}, \qquad b_2 = c + \frac{(d-c)\mu(\{d\})}{1-p}.$$

Then it holds that $b_1 \in (a, c \wedge \bar{\mu})$ and $b_2 \in (c \vee \bar{\mu}, d)$. Define the stopping time

$$\begin{aligned}\tau = \tau_{b_1,b_2}(\tau_{\bar{\mu}}) &+ \mathbb{1}_{\left\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_1\right\}} \inf\left\{t \in [0,\infty) \colon Y_{t+\tau_{b_1,b_2}(\tau_{\bar{\mu}})} \notin (a,c)\right\} \\ &+ \mathbb{1}_{\left\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_2\right\}} \inf\left\{t \in [0,\infty) \colon Y_{t+\tau_{b_1,b_2}(\tau_{\bar{\mu}})} \notin (c,d)\right\}.\end{aligned} \tag{III.4.5}$$

Figure III.4 depicts four realizations of a Brownian motion starting in $\bar{\mu}$ which is stopped at τ. Observe that Y_τ has distribution μ under $\mathbb{P}^y$. To see this, first note that

$$\begin{aligned}\mathbb{P}^y\left[Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})} = b_1\right] &= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\left\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_1\right\}} \,\middle|\, \mathcal{F}_{\tau_{\bar{\mu}}}\right]\right] \\ &= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\left\{Y_{\tau(\tau_{\bar{\mu}};b_1,b_2)}=b_1\right\}} \circ \theta_{\tau_{\bar{\mu}}} \,\middle|\, \mathcal{F}_{\tau_{\bar{\mu}}}\right]\right],\end{aligned}$$

where $\tau(\tau_{\bar{\mu}}; b_1, b_2) = \inf\left\{t \in [0,\infty) \colon Y_{\tau_{\bar{\mu}}+t} \notin (b_1,b_2)\right\} = \tau_{b_1,b_2}(\tau_{\bar{\mu}}) - \tau_{\bar{\mu}}$. Hence, the strong Markov property of Y and $\tau_{\bar{\mu}} = 0$, $\mathbb{P}^{\bar{\mu}}$-a.s., yield

$$\begin{aligned}\mathbb{P}^y\left[Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})} = b_1\right] &= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\left\{Y_{\tau(\tau_{\bar{\mu}};b_1,b_2)}=b_1\right\}} \circ \theta_{\tau_{\bar{\mu}}} \,\middle|\, \mathcal{F}_{\tau_{\bar{\mu}}}\right]\right] \\ &= \mathbb{E}^y\left[\mathbb{E}^{\bar{\mu}}\left[\mathbb{1}_{\left\{Y_{\tau(\tau_{\bar{\mu}};b_1,b_2)}=b_1\right\}}\right]\right] \\ &= \mathbb{E}^y\left[\mathbb{E}^{\bar{\mu}}\left[\mathbb{1}_{\left\{Y_{\tau_{b_1,b_2}(0)}=b_1\right\}}\right]\right] \\ &= \mathbb{P}^{\bar{\mu}}\left[Y_{\tau_{b_1,b_2}(0)} = b_1\right] \\ &= \frac{b_2 - \bar{\mu}}{b_2 - b_1} = p.\end{aligned}$$

Let

$$\begin{aligned}\tau(\tau_{\bar{\mu}}; b_1, b_2; a, c, d) = \tau - \tau_{b_1,b_2}(\tau_{\bar{\mu}}) = &\mathbb{1}_{\left\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_1\right\}} \inf\left\{t \in [0,\infty) \colon Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})+t} \notin (a,c)\right\} \\ &+ \mathbb{1}_{\left\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_2\right\}} \inf\left\{t \in [0,\infty) \colon Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})+t} \notin (c,d)\right\}.\end{aligned}$$

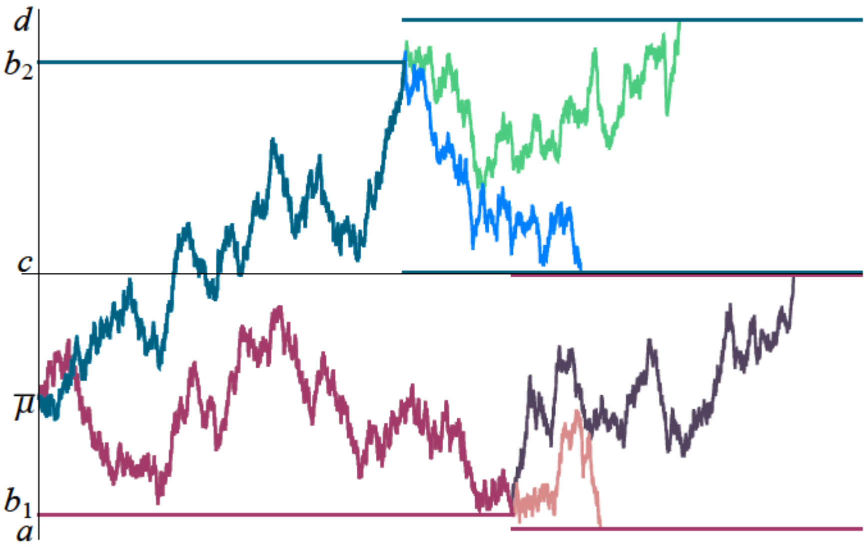


Figure III.4: Four realizations of a Brownian motion that starts in $\bar{\mu}$ and is stopped at the consecutive exit time τ, where τ is given by (III.4.5). Here $\mu = \frac{17}{24}\delta_{-1} + \frac{1}{12}\delta_0 + \frac{5}{24}\delta_1$, $b_1 = -\frac{17}{18}, b_2 = \frac{5}{6}$ and $p = \frac{3}{4}$. The barriers $a = -1$ and $c = 0$ as well as c and $d = 1$ enter at a different point in time which depends on the Brownian path.

Then the strong Markov property of Y implies

$$\begin{aligned}
\mathbb{P}^y[Y_\tau = a] &= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\{Y_\tau = a\}}\,\middle|\,\mathcal{F}_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}\right]\,\middle|\,\mathcal{F}_{\tau_{\bar{\mu}}}\right]\right] \\
&= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\{Y_{\tau(\tau_{\bar{\mu}};b_1,b_2;a,c,d)}=a\}}\circ\theta_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}\,\middle|\,\mathcal{F}_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}\right]\,\middle|\,\mathcal{F}_{\tau_{\bar{\mu}}}\right]\right] \\
&= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_1\}}\mathbb{1}_{\{Y_{\tau(\tau_{\bar{\mu}};b_1,b_2;a,c,d)}=a\}}\circ\theta_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}\,\middle|\,\mathcal{F}_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}\right]\,\middle|\,\mathcal{F}_{\tau_{\bar{\mu}}}\right]\right] \\
&= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_1\}}\mathbb{E}^{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}}\left[\mathbb{1}_{\{Y_{\tau(\tau_{\bar{\mu}};b_1,b_2;a,c,d)}=a\}}\right]\,\middle|\,\mathcal{F}_{\tau_{\bar{\mu}}}\right]\right] \\
&= \mathbb{E}^y\left[\mathbb{E}^y\left[\mathbb{1}_{\{Y_{\tau_{b_1,b_2}(\tau_{\bar{\mu}})}=b_1\}}\frac{c-b_1}{c-a}\,\middle|\,\mathcal{F}_{\tau_{\bar{\mu}}}\right]\right] \\
&= p\frac{c-b_1}{c-a} = \mu(\{a\}).
\end{aligned}$$

Similarly, one shows that $\mathbb{P}^y[Y_\tau = c] = \mu(\{c\})$ and $\mathbb{P}^y[Y_\tau = d] = \mu(\{d\})$ and hence, $Y_\tau \sim \mu$ under $\mathbb{P}^y$. For $p = \mu(\{a\})$ we obtain $b_1 = a$ and $b_2 = \frac{c\mu(\{c\})+d\mu(\{d\})}{1-\mu(\{a\})}$ and hence $\tau = \tau_2$, where τ_2 is the stopping time defined in (III.4.3) from the proof of Theorem III.1.5.
Allowing not only for left-hand side and right-hand side tangents in Algorithm III.3.6, we can construct τ as follows: Draw the tangent t_1 in c with slope $1 - 2p \in \big(2\mu(\{d\}) - 1, 1 - 2\mu(\{a\})\big)$. The intersection points of t_1 and u_{μ_0} are given by b_1 and b_2. Continue with the right-hand side tangent in a and the left-hand side tangent in d. The corresponding stopping time is given by (III.4.5).

The following example shows that in general a reduction to $\mathcal{A}_2(T, y)$, the set of probability measures in $\mathcal{A}(T, y)$ that are weighted sums of at most 2 Dirac measures, is not possible.

Example III.4.5. Let $(Y_t)_{t\in[0,\infty)}$ be a Brownian motion starting in 0 under $\mathbb{P}^0$ and let $f(x) = \mathbb{1}_{\{|x|\geq 1\}}$, $x \in \mathbb{R}$, be the payoff function. According to Remark III.1.2 the speed measure of Y is given by $m(dx) = 2\,dx$ and the function q_0 satisfies $q_0(x) = x^2$, $x \in \mathbb{R}$.

We claim that

$$V(T,0) = \sup_{\mu\in\mathcal{A}(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = T\wedge 1,$$

and

$$V_2(T,0) := \sup_{\mu\in\mathcal{A}_2(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = \begin{cases} \frac{T}{1+T}, & T<1, \\ 1, & T\geq 1. \end{cases} \tag{III.4.6}$$

To show this, first observe that the second constraint in the definition of $\mathcal{A}(T,0)$ ensures that all measures in $\mathcal{A}(T,0)$ are centered around 0. If $T\geq 1$, the measure μ^* given by

$$\mu^* = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$$

satisfies $\mu^*\in\mathcal{A}_2(T,0)\subseteq\mathcal{A}(T,0)$. Since f attains its maximum at -1 and 1 it follows that $V(T,0)=V_2(T,0)=\int_{\mathbb{R}} f(x)\mu^*(dx)=1$ in this case.

In the sequel assume that $T<1$. Observe that for every measure $\mu\in\mathcal{A}_2(T,0)$ at least one mass point is contained in $(-1,1)$. Due to the symmetry of the optimization problem in (III.4.6) and the form of f, we can restrict ourselves to measures of the form

$$\mu^S = \frac{1}{1+S}\delta_{-S} + \frac{S}{1+S}\delta_1 \in \mathcal{A}_2(T,0),$$

where $S\in(0,T]$. Then we obtain

$$V_2(T,0) = \sup_{\mu\in\mathcal{A}_2(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{S\in(0,T]} \int_{\mathbb{R}} f(x)\mu^S(dx) = \frac{T}{1+T}.$$

Theorem III.2.6 implies that in the maximization problem for V it is sufficient to consider measures $\mu\in\mathcal{A}_3(T,0)$. Moreover, since f is constant and maximal on $\mathbb{R}\backslash(-1,1)$ and symmetric, we can restrict ourselves to measures with mass points $-1, c$ and 1 for $c\in[0,1)$. The class of all centered probability measures $\mu^{S,c}$ with these three mass points and $\int_{\mathbb{R}} q_0(x)\mu^{S,c}(dx)=S$ is given by

$$\mu^{S,c} = \frac{S+c}{2(1+c)}\,\delta_{-1} + \frac{1-S}{1-c^2}\,\delta_c + \frac{S-c}{2(1-c)}\,\delta_1, \quad c\in[0,1), S\in[c,1].$$

Hence $\mu^{S,c}\in\mathcal{A}_3(T,0)$ if and only if $c\in[0,T]$ and $S\in[c,T]$. We have

$$\int_{\mathbb{R}} f(x)\mu^{S,c}(dx) = \frac{S-c^2}{1-c^2} = 1-\frac{1-S}{1-c^2},$$

which is maximized for $c=0$ and $S=T$. Hence we obtain,

$$V(T,0) = \sup_{\mu\in\mathcal{A}(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu\in\mathcal{A}_3(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = \int_{\mathbb{R}} f(x)\mu^*(dx) = T$$

with

$$\mu^* = \frac{T}{2}\,\delta_{-1} + (1-T)\,\delta_0 + \frac{T}{2}\,\delta_1.$$

The proof of Theorem III.1.5 yields that the corresponding optimal stopping time is given by

$$\tau = \tau_{-1,b}(0) + \mathbb{1}_{\{Y_{\tau_{-1,b}(0)}=b\}} \inf\left\{t\in[0,\infty)\colon Y_{t+\tau_{-1,b}(0)}\in\{0,1\}\right\},$$

where $b=\frac{T}{2-T}$.

Example III.4.6. The framework of Section III.1 allows to solve stopping problems where the process to stop is not necessarily characterized as solution of an SDE. One such example is Brownian motion on $\mathbb{R}$ sticky at 0. This process evolves like a Brownian motion outside 0 but spends a Lebesgue-positive amount of time at zero without having intervals of zeros. More formally, let Y be a general diffusion in natural scale with state space $J = \mathbb{R}$ and speed measure

$$m(dx) = 2dx + 2\kappa\delta_0(dx),$$

where $\kappa \in [0, \infty)$. It follows that the function q_0 satisfies

$$q_0(x) = x^2 + \kappa|x|, \quad \forall x \in \mathbb{R}.$$

One can generalize the results of Example III.4.5 to the sticky case. Let $f(x) = \mathbb{1}_{\{|x|\geq 1\}}$, $x \in \mathbb{R}$, be the payoff function. Then similar calculations as in Example III.4.5 show that

$$V(T,0) = \sup_{\mu \in \mathcal{A}(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = \frac{T}{1+\kappa} \wedge 1,$$

with optimal measure

$$\mu^* = \frac{1}{2}\left(\frac{T}{1+\kappa} \wedge 1\right)\delta_{-1} + \left(1 - \frac{T}{1+\kappa}\right)^+ \delta_0 + \frac{1}{2}\left(\frac{T}{1+\kappa} \wedge 1\right)\delta_1.$$

Moreover, a straight-forward calculation shows that for $T < 1+\kappa$ the supremum over $\mathcal{A}_2(T,0)$ is strictly smaller than the value function $V(T,0)$. Indeed, for $\kappa > 0$ it holds that

$$\sup_{\mu \in \mathcal{A}_2(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = \begin{cases} \frac{1+2\kappa+T-\sqrt{4T+(2\kappa+1-T)^2}}{4\kappa}, & T < 1+\kappa, \\ 1, & T \geq 1+\kappa. \end{cases}$$

For $\kappa = 0$ we are in the setting of Example III.4.5 and thus

$$\sup_{\mu \in \mathcal{A}_2(T,0)} \int_{\mathbb{R}} f(x)\mu(dx) = \begin{cases} \frac{T}{T+1} & T < 1, \\ 1, & T \geq 1. \end{cases}$$

The parameter κ controls the amount of time spent at zero by the sticky Brownian motion. For large values of κ, the process is held longer in zero, and the optimal value $V(T,0)$, which coincides with the probability of stopping the process Y outside the interval $(-1,1)$, is small.

Next we revisit Example II.4.9. The optimal stopping time τ_ϑ associated to the optimal control α^* is not a first exit time of an interval. Observe that for T sufficiently large the law of the Brownian motion at τ_ϑ has support $\mathbb{R}\backslash\big((-1,0) \cup (0,1)\big)$.

Example III.4.7. Let $Y = W$ be a Brownian motion starting in y under $\mathbb{P}^y$ and let $f(x) = x^2\mathbb{1}_{\{|x|\geq 1\}}$. In Example II.4.9 we have proven that

$$V(T,y) = \begin{cases} \frac{T}{T+(1-|y|)^2}, & T < |y|(1-|y|), \\ T + y^2, & T \geq |y|(1-|y|). \end{cases}$$

Observe that for $T \geq 1 - y^2$ an optimal stopping time is given by $\tau(a,b)$, where $a \leq -1 \wedge y$, $b \geq 1 \vee y$ and $(b-y)(y-a) = T$. For $|y| < 1$ and $T \leq |y|(1-|y|)$ the first exit time of $\left(x - \frac{T}{1-x}, 1\right)$ if $y \in (0,1)$ and of the interval $\left(-1, x + \frac{T}{1+x}\right)$ if $y \in (-1,0)$ is optimal. These stopping times correspond to the optimal control α^*. The law of the Brownian motion at these stopping times is purely atomic and has exactly two mass points.

Let $|y| < 1$ and $T \in \big(|y|(1-|y|), 1-y^2\big)$. Denote by τ_ϑ the stopping time associated to the optimal control $\alpha_s^* = \big(-2Y_s + \operatorname{sgn}(Y_\vartheta)\big)\mathbb{1}_{\{s\geq\vartheta\}}$, where $\vartheta = \inf\{t \in [0,\infty) \colon T - t < |Y_t|(1-|Y_t|)\}$.

Then the law of W_{τ_ϑ} has support in $\mathbb{R}\backslash\big((-1,0)\cup(0,1)\big)$. There exists another stopping time τ that is optimal for V with $\mu^* := \mathrm{Law}(W_\tau) \in \mathcal{A}_3(T,y)$ and

$$\mu^* = \frac{1}{2}\big(T+y^2-y\big)\delta_{-1} + \big(1-T-y^2\big)\delta_0 + \frac{1}{2}\big(T+y^2+y\big)\delta_1.$$

Moreover, similar to Example III.4.5 one can show that μ^* is optimal for $V(T,y)$ and

$$\sup_{\mu\in\mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \frac{T}{T+(1-|y|^2)} < V(T,y).$$

III.5 Linear Optimization and Extreme Points

Now we extend Theorem III.2.6 to payoff functions satisfying Assumption (**B**) instead of the stronger Assumptions **(B1)** and **(B2)**. The linear nature of the measure optimization problem (III.2.4) allows to conclude that the maximum values are attained by extreme points that turn out to be weighted sums of at most three Dirac measures.

Corollary III.2.5 reveals that the optimal stopping problem (III.1.3) is equivalent to a measure optimization problem over the set $\mathcal{A}(T,y)$. Notice that the functional $\mu \mapsto \int_{\mathbb{R}} f(x)\mu(dx)$ is linear on $\mathcal{A}(T,y)$. We have thus obtained a linear problem over a set of probability measures μ with some integrability constraints. Recall linear optimization problems in $\mathbb{R}^n$ with linear constraints: If the set A of points in $\mathbb{R}^n$ satisfying the linear constraints is non-empty and contains at least one extreme point (e.g. if A is compact), then the maximum value is attained by extreme points of A, see e.g. Theorem 2.7, 2.8 and Corollary 2.3 in [10]. We have a similar result for an optimization problem $\int_{\mathbb{R}} g(x)\mu(dx)$ over measures $\mu \in \mathcal{M}$ satisfying moment constraints of the form $\int_{\mathbb{R}} f_i(x)\mu(dx) \le c_i$, where g and f_i are measurable, $c_i \in \mathbb{R}$, $1 \le i \le n$. The maximum value of $\int_{\mathbb{R}} g(x)\mu(dx)$ is also attained in the set of extreme points, see [71]. Furthermore, the extreme points are contained in the set of all weighted Dirac measures with at most $n+1$ mass points satisfying the moment constraints.

In the following we denote the extreme points of a convex set $A \subseteq \mathcal{M}$ by $\mathcal{E}(A)$ and for any $M \subseteq \mathcal{M}$ we denote by M_3 the set of all measures in M which are a weighted sum of at most 3 Dirac measures.

Now we prove Theorem III.2.6, which allows us to reduce the optimization problem (III.2.4) to an optimization problem over the set $\mathcal{A}_3(T,y)$, i.e.

$$V(T,y) = \sup_{\mu\in\mathcal{A}(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu\in\mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

Proof of Theorem III.2.6. Since $\mathcal{A}_3(T,y) \subseteq \mathcal{A}(T,y)$ we conclude that

$$V(T,y) \ge \sup_{\mu\in\mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

For the reverse inequality we consider two cases. In the first case we assume that all measures μ in $\mathcal{A}(T,y)$ are centered around y, i.e. $\bar{\mu} = y$. Observe that $\bar{\mu} = y$ for all $\mu \in \mathcal{A}(T,y)$ if and only if one of the following four cases is satisfied:

1. J is bounded,
2. $l > -\infty$, $r = \infty$ and $m\big((y,\infty)\big) = \infty$,
3. $l = -\infty, r < \infty$ and $m\big((-\infty,y)\big) = \infty$,
4. $J = \mathbb{R}$, $m\big((y,\infty)\big) = \infty$ and $m\big((-\infty,y)\big) = \infty$.

The optimization problem (III.2.4) can be rewritten as

$$V(T,y) = \sup_{t\in[0,T]} \sup_{\mu\in\mathcal{D}(t,y)} \int_{\mathbb{R}} f(x)\mu(dx),$$

where $\mathcal{D}(t,y) = \{\mu \in \mathcal{M}^1 \colon \bar{\mu} = y \text{ and } \int_{\mathbb{R}} q_y(x)\mu(dx) = t\}$, $0 \le t \le T$. Proposition 3.1 and Theorem 3.2 in [71] imply that

$$\sup_{\mu\in\mathcal{D}(t,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu\in\mathcal{E}(\mathcal{D}(t,y))} \int_{\mathbb{R}} f(x)\mu(dx).$$

From Theorem 2.1(b) in [71] we conclude that

$$\mathcal{E}\big(\mathcal{D}(t,y)\big) = \Bigg\{\mu \in \mathcal{D}(t,y) : \mu = \sum_{j=1}^{m} p_j\delta_{a_j},\ p_j > 0,\ \sum_{j=1}^{m} p_j = 1,\ a_j \in J,$$
$$\big\{\big(1, a_j, q_y(a_j)\big)\big\}_{1\le j\le m} \text{ is linearly independent, } 1 \le m \le 3\Bigg\}.$$

Let $a_1, a_2, a_3 \in J$. Assume without loss of generality that $a_1 < a_2 < a_3$. Let $\lambda_i \in \mathbb{R}$, $1 \le i \le 3$, such that

$$\begin{pmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \\ q_y(a_1) & q_y(a_2) & q_y(a_3) \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

In particular, we have $\lambda_2 = -\frac{a_3-a_1}{a_3-a_2}\lambda_1$, $\lambda_3 = \frac{a_2-a_1}{a_3-a_2}\lambda_1$ and

$$\lambda_1 \frac{a_3 - a_1}{a_3 - a_2} \left(\frac{a_3 - a_2}{a_3 - a_1} q_y(a_1) + \frac{a_2 - a_1}{a_3 - a_1} q_y(a_3) - q_y(a_2) \right) = 0.$$

Since q_y is strictly convex, it follows that $\lambda_1 = 0$ and hence, $\lambda_2 = \lambda_3 = 0$. Thus, the vectors $\big\{\big(1, a_j, q_y(a_j)\big)\big\}_{1\le j\le m}$, $1 \le m \le 3$, are linearly independent. As a consequence, $\mathcal{E}\big(\mathcal{D}(t,y)\big) = \mathcal{D}_3(t,y)$ and hence,

$$\sup_{\mu\in\mathcal{D}(t,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu\in\mathcal{D}_3(t,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

For $t \in [0,T]$ it holds that $\mathcal{D}_3(t,y) \subseteq \mathcal{A}_3(T,y)$. To sum up, we have

$$\begin{aligned} V(T,y) = \sup_{t\in[0,T]} \sup_{\mu\in\mathcal{D}_3(t,y)} \int_{\mathbb{R}} f(x)\mu(dx) &= \sup_{\mu\in\mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) \\ &\le \sup_{\mu\in\mathcal{A}(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = V(T,y). \end{aligned}$$

This proves Theorem III.1.5 in the first case. In the second case the set $\mathcal{A}(T,y)$ also contains measures which are not centered around y. We define

$$\mathcal{A}^+(T,y) = \begin{cases} \{\mu \in \mathcal{A}(T,y) \colon \bar{\mu} \ge y\}, & \text{if } \exists\, \mu \in \mathcal{A}(T,y) \text{ with } \bar{\mu} > y, \\ \emptyset, & \text{if } \bar{\mu} \le y \text{ for all } \mu \in \mathcal{A}(t,y), \end{cases}$$

$$\mathcal{A}^-(T,y) = \begin{cases} \{\mu \in \mathcal{A}(T,y) \colon \bar{\mu} \le y\}, & \text{if } \exists\, \mu \in \mathcal{A}(T,y) \text{ with } \bar{\mu} < y, \\ \emptyset, & \text{if } \bar{\mu} \ge y \text{ for all } \mu \in \mathcal{A}(t,y). \end{cases}$$

Observe that at least one of the sets $\mathcal{A}^+(T,y)$ or $\mathcal{A}^-(T,y)$ is non-empty and that (III.2.4) can be reduced to the two optimization problems

$$\sup_{\mu\in\mathcal{A}^+(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) \quad \text{and} \quad \sup_{\mu\in\mathcal{A}^-(T,y)} \int_{\mathbb{R}} f(x)\mu(dx),$$

where we follow the convention that the supremum over the empty set is equal to $-\infty$. If $\mathcal{A}^+(T,y)$ is non-empty, then

$$\begin{aligned}\mathcal{A}^+(T,y) &= \left\{\mu \in \mathcal{M}^1 \colon \bar{\mu} \geq y, \int_{\mathbb{R}} q_y(x)\mu(dx) \leq T - H(y,\bar{\mu})\right\} \\ &= \left\{\mu \in \mathcal{M}^1 \colon \int_{\mathbb{R}} -x\,\mu(dx) \leq -y, \int_{\mathbb{R}} \big(q_y(x) + Cx\big)\mu(dx) \leq T + Cy\right\},\end{aligned}$$

where $C = m\big((-\infty,y)\big) + \frac{1}{2}m\big(\{y\}\big) < \infty$ (see Remark III.2.1). Therefore, Proposition 3.1 and Theorem 3.2 in [71] imply that

$$\sup_{\mu \in \mathcal{A}^+(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu \in \mathcal{E}(\mathcal{A}^+(T,y))} \int_{\mathbb{R}} f(x)\mu(dx).$$

By Theorem 2.1(a) in [71] we have $\mathcal{E}\big(A^+(T,y)\big) \subseteq \mathcal{A}_3^+(T,y)$. Thus,

$$\sup_{\mu \in \mathcal{A}^+(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{\mu \in \mathcal{A}_3^+(T,y)} \int_{\mathbb{R}} f(x)\mu(dx). \tag{III.5.1}$$

If $\mathcal{A}^-(T,y)$ is non-empty, similar arguments show that (III.5.1) holds with $\mathcal{A}^+(T,y)$ and $\mathcal{A}_3^+(T,y)$ replaced by $\mathcal{A}^-(T,y)$ and $\mathcal{A}_3^-(T,y)$, respectively. Since $\mathcal{A}_3(T,y) = \mathcal{A}_3^-(T,y) \cup \mathcal{A}_3^+(T,y)$ we conclude that

$$V(T,y) = \sup_{\mu \in \mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

□

III.6 Existence of an Optimizer

In this section we first show that the supremum in (III.1.3) is not always attained. We then state sufficient conditions guaranteeing that the supremum in (III.2.5) and thus in (III.1.3) is attained.

Example III.6.1. Let $f_1(x) = x^2 \frac{|x|}{1+|x|}$, $x \in \mathbb{R}$, and Y be a Brownian motion starting in 0 under $\mathbb{P}^0$. In this case there does not exist an optimal stopping time for $T \in (0,\infty)$. To prove this, let $V(T) := \sup_{\tau \in \mathcal{S}(T,0)} \mathbb{E}^0[f_1(Y_\tau)]$, $T \in (0,\infty)$. Moreover, consider the second payoff function $f_2(x) = x^2$. Note that for any integrable stopping time τ we have $\mathbb{E}^0[Y_\tau^2] = \mathbb{E}^0[\tau]$, see (II.4.8). Therefore, $\widetilde{V}(T) := \sup_{\tau \in \mathcal{S}(T,0)} \mathbb{E}^0[f_2(Y_\tau)] = T$.

One can show that $V = \widetilde{V}$. Indeed, on the one hand it must hold that $V \leq \widetilde{V}$ since $f_1 \leq f_2$. On the other hand, for the stopping times $\tau_n = \tau_{-1/n,nT}$ we have $\mathbb{E}^0[\tau_n] = T$ and

$$\mathbb{E}^0\left[f_1(Y_{\tau_n})\right] = \frac{nT}{\frac{1}{n} + nT}\frac{1}{n^2}\frac{\frac{1}{n}}{1+\frac{1}{n}} + \frac{\frac{1}{n}}{\frac{1}{n}+nT}n^2T^2\frac{nT}{1+nT} \longrightarrow T,$$

as $n \to \infty$, and hence $V \geq \widetilde{V}$.

From $V_1 = V_2$ we can deduce that the supremum cannot be attained in V_1, because for any stopping time $\tau \neq 0$ with $\mathbb{E}^0[\tau] < \infty$ we have $\mathbb{P}^0[f_1(Y_\tau) < f_2(Y_\tau)] > 0$.

We now establish the existence of an optimal measure in $\mathcal{A}_3(T,y)$ in (III.2.5) under mild conditions on the payoff function f.

Theorem III.6.2. *Assume that $f \colon J \to \mathbb{R}$ is upper semi-continuous with*

$$\limsup_{x \uparrow r} \frac{f(x)}{q_y(x)} \leq 0, \ \text{if } r \notin J \quad \text{and} \quad \limsup_{x \downarrow l} \frac{f(x)}{q_y(x)} \leq 0, \ \text{if } l \notin J.$$

Then there exists an optimal measure in $\mathcal{A}_3(T,y)$ for (III.2.5) *and an optimal stopping time in $\mathcal{S}_3(T,y)$ for* (III.1.5).

Proof. Throughout the proof we denote by $C_+, C_- \in [0,\infty]$ the extended real numbers given by $C_+ = m\big((y,\infty)\big) + \frac{1}{2}m\big(\{y\}\big)$ and $C_- = m\big((-\infty,y)\big) + \frac{1}{2}m\big(\{y\}\big)$. Let

$$\mu_n = \sum_{j=1}^{3} p_n^j \delta_{x_n^j} \in \mathcal{A}(T,y), \quad n \in \mathbb{N},$$

be a sequence of measures such that

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\mu_n(dx) = V(T,y).$$

If the sequence $\big(x_n^1\big)_{n\in\mathbb{N}}$ is unbounded, choose a subsequence, also denoted by $\big(x_n^1\big)_{n\in\mathbb{N}}$, such that either $\lim_{n\to\infty} x_n^1 = -\infty =: x^1$ or $\lim_{n\to\infty} x_n^1 = \infty =: x^1$. If $\big(x_n^1\big)_{n\in\mathbb{N}}$ is bounded, extract a subsequence such that $\lim_{n\to\infty} x_n^1 = x^1 \in \bar{J}$. Here $\bar{J}$ denotes the closure of J in $\mathbb{R}$ with respect to the Euclidean metric. By extracting further subsequences, proceed in the same way with $\big(x_n^2\big)_{n\in\mathbb{N}}$ and $\big(x_n^3\big)_{n\in\mathbb{N}}$. Refine once again the sequence to obtain that $\big(p_n^1, p_n^2, p_n^3\big) \to \big(p^1,p^2,p^3\big) \in [0,1]^3$ as $n\to\infty$. Overall, we have for $n \to \infty$ that

$$\big(x_n^1, x_n^2, x_n^3, p_n^1, p_n^2, p_n^3\big) \to \big(x^1,x^2,x^3,p^1,p^2,p^3\big) \in \big(\bar{J} \cup \{-\infty\} \cup \{\infty\}\big)^3 \times [0,1]^3.$$

Note that $x^j = \infty$ and $x^j = -\infty$, $j \in \{1,2,3\}$, is only possible if $r = \infty$ and $l = -\infty$, respectively.

Let

$$\begin{aligned} K &= \big\{j \in \{1,2,3\} \colon x^j \in J\big\}, \\ \bar{K}^+ &= \big\{j \in \{1,2,3\} \colon x^j \notin J,\, x^j = r\big\}, \\ \bar{K}^- &= \big\{j \in \{1,2,3\} \colon x^j \notin J,\, x^j = l\big\}. \end{aligned}$$

Define $\mu = \sum_{k\in K} p^k \delta_{x^k}$. We show that μ is an optimizer for (III.2.5).

From the fact that for all $i \in \{1,2,3\}$ it holds

$$0 \le p_n^i q_y(x_n^i) \le \sum_{j=1}^{3} p_n^j q_y(x_n^j) = \int_{\mathbb{R}} q_y(x)\mu_n(dx) \le T \tag{III.6.1}$$

and that $\lim_{n\to\infty} q_y(x_n^i) = \infty$ for all $i \in \{1,2,3\} \setminus K$ by Lemma III.1.1, it follows for all $i \in \{1,2,3\} \setminus K$ that

$$\lim_{n\to\infty} p_n^i = \lim_{n\to\infty} \frac{1}{q_y(x_n^i)} p_n^i q_y(x_n^i) = 0. \tag{III.6.2}$$

We conclude from (III.6.2) that

$$\mu(J) = \sum_{k\in K} p^k = \lim_{n\to\infty} \sum_{k\in K} p_n^k = \lim_{n\to\infty} \sum_{j=1}^{3} p_n^j = \lim_{n\to\infty} \mu_n(J) = 1.$$

Thus, $\mu \in \mathcal{M}^1$. Next we show that $\mu \in \mathcal{A}_3(T,y)$. To this end we distinguish four cases.

1. $\boxed{l > -\infty,\ r < \infty:}$ Observe that in this case we have $\overline{\mu_n} = y$ for all $n \in \mathbb{N}$. This together with (III.6.2) ensures that

$$y = \lim_{n\to\infty} \overline{\mu_n} = \lim_{n\to\infty} \left(\sum_{k\in K} p_n^k x_n^k + \sum_{i\in\{1,2,3\}\setminus K} p_n^i x_n^i \right) = \sum_{k\in K} p^k x^k = \overline{\mu}.$$

Moreover, continuity and non-negativity of q_y on J imply that

$$\int_{\mathbb{R}} q_y(x)\mu(dx) = \lim_{n\to\infty} \sum_{k\in K} p_n^k q_y\big(x_n^k\big) \le \limsup_{n\to\infty} \sum_{j=1}^{3} p_n^j q_y\big(x_n^j\big) \le T = T - H(y,\overline{\mu}). \tag{III.6.3}$$

This proves that $\mu \in \mathcal{A}_3(T,y)$.

2. $\boxed{l > -\infty,\ r = \infty:}$ In this case we know that $\overline{\mu_n} \le y$ for all $n \in \mathbb{N}$. Let us first assume that $m\big((y,\infty)\big) = \infty$. Then it holds that $\overline{\mu_n} = y$ for all $n \in \mathbb{N}$. Moreover, we have for all $i \in \bar{K}^+$ that $\lim_{n\to\infty} \frac{q_y(x_n^i)}{x_n^i} = C_+ = \frac{1}{2}m\big(\{y\}\big) + m\big((y,\infty)\big) = \infty$ and hence, with (III.6.1),

$$\lim_{n\to\infty} p_n^i x_n^i = \lim_{n\to\infty} p_n^i q_y(x_n^i)\frac{x_n^i}{q_y(x_n^i)} = 0. \tag{III.6.4}$$

This and (III.6.2) show that

$$y = \lim_{n\to\infty} \overline{\mu_n} = \lim_{n\to\infty}\left(\sum_{k\in K} p_n^k x_n^k + \sum_{i\in\bar{K}^+} p_n^i x_n^i + \sum_{i\in\bar{K}^-} p_n^i x_n^i\right) = \sum_{k\in K} p^k x^k = \overline{\mu}.$$

Then the same reasoning as in (III.6.3) demonstrates that $\int_{\mathbb{R}} q_y(x)\mu_n(dx) \le T - H(y,\overline{\mu})$ and hence $\mu \in \mathcal{A}_3(T,y)$.

Let us now assume that $m\big((y,\infty)\big) < \infty$. Equation (III.6.2) implies that

$$\begin{aligned} y &\ge \limsup_{n\to\infty} \overline{\mu_n} = \limsup_{n\to\infty}\left(\sum_{k\in K} p_n^k x_n^k + \sum_{i\in\bar{K}^+} p_n^i x_n^i + \sum_{i\in\bar{K}^-} p_n^i x_n^i\right) \\ &\ge \sum_{k\in K} p^k x^k + \limsup_{n\to\infty} \sum_{i\in\bar{K}^-} p_n^i x_n^i = \overline{\mu}. \end{aligned} \tag{III.6.5}$$

Moreover, it holds for all $n \in \mathbb{N}$ that

$$\begin{aligned} T &\ge \int_{\mathbb{R}} q_y(x)\mu_n(dx) + H(y,\overline{\mu_n}) = \sum_{j=1}^{3} p_n^j q_y\big(x_n^j\big) + C_+\left(y - \sum_{j=1}^{3} p_n^j x_n^j\right) \\ &= \sum_{j=1}^{3} p_n^j\Big(q_y\big(x_n^j\big) - C_+\big(x_n^j - y\big)\Big). \end{aligned} \tag{III.6.6}$$

It follows with (III.6.5) that

$$\begin{aligned} \lim_{n\to\infty} \sum_{k\in K} p_n^k\left(q_y\big(x_n^k\big) - C_+\big(x_n^k - y\big)\right) &= \sum_{k\in K} p^k\left(q_y\big(x^k\big) - C_+\big(x^k - y\big)\right) \\ &= \int_{\mathbb{R}} q_y(x)\mu(dx) + H(y,\bar{\mu}). \end{aligned} \tag{III.6.7}$$

Combining (III.6.1) and $\lim_{x\to\infty}\frac{q_y(x)}{x-y} = C_+$ yields that

$$\lim_{n\to\infty} \sum_{i\in\bar{K}^+} p_n^i\left(q_y\big(x_n^i\big) - C_+\big(x_n^i - y\big)\right) = 0. \tag{III.6.8}$$

Moreover, the non-negativity of q_y and (III.6.2) imply that

$$\liminf_{n\to\infty} \sum_{i\in\bar{K}^-} p_n^i\left(q_y\big(x_n^i\big) - C_+\big(x_n^i - y\big)\right) \ge \liminf_{n\to\infty} \sum_{i\in\bar{K}^-} C_+ p_n^i\big(y - x_n^i\big) = 0. \tag{III.6.9}$$

Combining (III.6.6), (III.6.7), (III.6.8) and (III.6.9) proves that $\mu \in \mathcal{A}_3(T,y)$.

3. $\boxed{l = -\infty,\ r < \infty:}$ This case is analog to the case $l > -\infty$, $r = \infty$.

4. $\boxed{l = -\infty,\ r = \infty:}$ In this case no conditions on $\bar{\mu}$ have to be verified. Assume first that $m\big((y,\infty)\big) = \infty$ and $m\big((-\infty,y)\big) = \infty$. As in (III.6.4) it follows in this case that $\lim_{n\to\infty} p_n^i x_n^i = 0$

for all $i \in \bar{K}^+ \cup \bar{K}^-$. In addition, it holds that $\overline{\mu_n} = y$ for all $n \in \mathbb{N}$. Hence, we conclude that $y = \lim_{n\to\infty} \overline{\mu_n} = \overline{\mu}$. As in (III.6.3) we obtain that $\mu \in \mathcal{A}_3(T, y)$. Next, assume that $m\big((y, \infty)\big) < \infty$ and $m\big((-\infty, y)\big) = \infty$. Then it holds that $y \geq \overline{\mu_n}$. Furthermore, we obtain as in (III.6.4) that $\lim_{n\to\infty} p_n^i x_n^i = 0$ for all $i \in \bar{K}^-$. This, together with the fact that $y \geq \overline{\mu_n}$ for all $n \in \mathbb{N}$, proves that $y \geq \overline{\mu}$ (see also (III.6.5)). Since $\lim_{n\to\infty} x_n^i = -\infty$ for all $i \in \bar{K}^-$ we conclude that

$$\liminf_{n\to\infty} \sum_{i\in\bar{K}^-} p_n^i \left(q_y\big(x_n^i\big) - C_+\big(x_n^i - y\big)\right) \geq \liminf_{n\to\infty} \sum_{i\in\bar{K}^-} C_+ p_n^i \big(y - x_n^i\big) \geq 0. \tag{III.6.10}$$

Then proceeding exactly as in (III.6.6), (III.6.7) and (III.6.8) shows that $\mu \in \mathcal{A}_3(T, y)$. The case $m\big((y, \infty)\big) = \infty$ and $m\big((-\infty, y)\big) < \infty$ can be treated analogously. Finally, we assume that $m\big((y, \infty)\big) < \infty$ and $m\big((-\infty, y)\big) < \infty$. Without loss of generality we also assume that $\overline{\mu} \leq y$. In this case we obtain for all $n \in \mathbb{N}$ that

$$\begin{aligned} T &\geq \int_{\mathbb{R}} q_y(x)\mu_n(dx) + H(y, \overline{\mu_n}) \\ &= \sum_{j=1}^{3} p_n^j q_y\big(x_n^j\big) + C_+ \left(y - \sum_{j=1}^{3} p_n^j x_n^j\right) + \mathbb{1}_{(y,\infty)}(\overline{\mu_n})(C_+ + C_-)\left(\overline{\mu_n} - y\right) \\ &\geq \sum_{j=1}^{3} p_n^j \Big(q_y\big(x_n^j\big) - C_+\big(x_n^j - y\big)\Big). \end{aligned}$$

Proceeding as in (III.6.7), (III.6.8) and (III.6.10) proves that $\mu \in \mathcal{A}_3(T, y)$.

To summarize we have shown that $\mu \in \mathcal{A}_3(T, y)$ in any possible case. It remains to show the optimality of μ. First note that $V(T, y) \geq \int_{\mathbb{R}} f(x)\mu(dx)$. The assumptions $\limsup_{x\uparrow r} \frac{f(x)}{q_y(x)} \leq 0$ if $r \notin J$ and $\limsup_{x\downarrow l} \frac{f(x)}{q_y(x)} \leq 0$ if $l \notin J$ together with (III.6.1) imply that for $i \in \bar{K}^- \cup \bar{K}^+$

$$\limsup_{n\to\infty} p_n^i f\big(x_n^i\big) = \limsup_{n\to\infty} p_n^i q_y\big(x_n^i\big) \frac{f\big(x_n^i\big)}{q_y\big(x_n^i\big)} \leq 0. \tag{III.6.11}$$

Finally, the upper semi-continuity of f and (III.6.11) result in

$$\begin{aligned} \int_{\mathbb{R}} f(x)\mu(dx) &= \sum_{k\in K} p^k f(x^k) \geq \limsup_{n\to\infty} \sum_{k\in K} p_n^k f(x_n^k) \\ &\geq \limsup_{n\to\infty} \sum_{k\in K} p_n^k f\big(x_n^k\big) + \limsup_{n\to\infty} \sum_{i\in\bar{K}^-\cup\bar{K}^+} p_n^i f\big(x_n^i\big) \\ &\geq \limsup_{n\to\infty} \sum_{j=1}^{3} p_n^j f\big(x_n^j\big) = \lim_{n\to\infty} \int_{\mathbb{R}} f(x)\mu_n(dx) = V(T, y). \end{aligned}$$

Therefore, we conclude that $V(T, y) = \int_{\mathbb{R}} f(x)\mu(dx)$. Finally, the proof of Theorem III.1.5 allows to construct a stopping time in $\mathcal{S}_3(T, y)$ which is optimal in (III.1.5). □

Remark III.6.3. Example III.6.1 shows that the condition that $\limsup_{x\uparrow r} \frac{f(x)}{q_y(x)} \leq 0$ if $r \notin J$ and $\limsup_{x\downarrow l} \frac{f(x)}{q_y(x)} \leq 0$ if $l \notin J$ in Theorem III.6.2 cannot be weakened in general.

III.7 General Diffusions

In this section we show how to deal with the optimal stopping problem (III.1.3) if Y is not in natural scale. Throughout, we suppose that Y satisfies all the properties of Section III.1 apart from being in natural scale.

Let s be the scale function of Y and m its speed measure. Define $Z_t = s(Y_t)$, $t \in [0,\infty)$. Then Z (or more precisely $(Q^z)_{z\in s(J)}$, where Q^z is the $\mathbb{P}^{s^{-1}(z)}$ distribution of $\big(s(Y_t)\big)_{t\in[0,\infty)}$) is a diffusion in natural scale on $J^Z := s(J)$ by Theorem 46.12, in [59, Chapter V] and Theorem 2.1 in [15]. The speed measure m^Z of Z is given by $m^Z = m \circ s^{-1}$, i.e. $m^Z\big((a,b)\big) = m\big((s^{-1}(a), s^{-1}(b))\big)$ for all $[a,b] \subseteq J^Z$. Denote by $(l^Z, r^Z) := \big(s(l), s(r)\big)$ the interior of J^Z and for $z \in \big(l^Z, r^Z\big)$ let q_z^Z be the function on $\overline{J^Z}$ defined in (III.1.1) with m replaced by m^Z. Recall that we assume that $y \in (l,r)$. In this section we impose the following assumption on $f\colon J \to \mathbb{R}$.

Assumption. There exists $C(y) \in [0,\infty)$ such that

$$f(w) \geq -C(y)\left(1 + q^Z_{s(y)}\big(s(w)\big)\right), \quad \forall w \in J. \tag{III.7.1}$$

In terms of the scale function and the speed measure of the process Y the condition (III.7.1) reads as follows

$$f(w) \geq -C(y)\left(1 + \frac{1}{2}m\big(\{y\}\big)|s(w)-s(y)| + \int_{s(y)}^{s(w)} m\big((y, s^{-1}(u))\big)du\right), \quad w \in J.$$

The optimal stopping problem (III.1.3) can be rewritten as

$$V(T,y) = \sup_{\tau\in\mathcal{S}(T,y)} \mathbb{E}^{\mathbb{P}^y}[f(Y_\tau)] = \sup_{\tau\in\mathcal{S}^Z(T,s(y))} \mathbb{E}^{Q^{s(y)}}\big[(f\circ s^{-1})(Z_\tau)\big],$$

where $\mathcal{S}^Z(T,z)$ denotes the set of all $(\mathcal{F}_t)$-stopping times with $\mathbb{E}^{Q^z}[\tau] \leq T$. Similar to Section II.3.1, we can convert the optimal stopping problem with reward function f for the process Y under $\mathbb{P}^y$ into an optimal stopping problem with reward function $f\circ s^{-1}$ for Z under $Q^{s(y)}$. Observe that (III.7.1) (like Assumption (**B**) in Section III.1) entails that the expectation $\mathbb{E}^{Q^{s(y)}}\big[(f\circ s^{-1})(Z_\tau)\big]$ exists for all $\tau \in \mathcal{S}^Z\big(T, s(y)\big)$.

For $z \in \big(l^Z, r^Z\big)$ let $\mathcal{A}^Z(T,z)$ be the set of all probability measures ν on $\mathbb{R}$ with support in J^Z and finite first moment satisfying Properties 1 and 2 of Section III.2, where $l = l^Z$, $r = r^Z$, $q_z = q_z^Z$ and $m = m^Z$. Denote by $\mathcal{A}_3^Z(T,z)$ the set of all probability measures in $\mathcal{A}^Z(T,z)$ which can be written as a weighted sum of at most 3 Dirac measures. Theorem III.2.6 implies that

$$V(T,y) = \sup_{\nu\in\mathcal{A}_3^Z(T,s(y))} \int_{\mathbb{R}} f\big(s^{-1}(x)\big)\,\nu(dx). \tag{III.7.2}$$

In the reduced optimization problem (III.7.2) the set $\mathcal{A}_3^Z\big(T,s(y)\big)$ depends on the process Z, its state space J^Z and speed measure m^Z. Next we aim at characterizing the set of measure $\mathcal{A}_3^Z\big(T,s(y)\big)$ in terms of the primal process Y, its state space J and speed measure m. First observe that an $(\mathcal{F}_t)$-stopping time τ embeds μ in Y under $\mathbb{P}^y$ if and only if τ embeds $\nu := \mu\circ s^{-1}$ in Z under $Q^{s(y)}$. In order to transfer the properties of $\nu \in \mathcal{A}_3^Z\big(T,s(y)\big)$ to μ we introduce the set $\mathcal{A}^s(T,y)$ of probability measures ρ on $\mathbb{R}$ with support in J which satisfy the following properties:

1. $\int_J |s(x)|\rho(dx) < \infty$.
2. a) If $s(l) > -\infty$, then $\int_J s(x)\,\rho(dx) \leq s(y)$.
   b) If $s(r) < \infty$, then $\int_J s(x)\,\rho(dx) \geq s(y)$.
3. ρ satisfies the following integrability condition

$$\begin{aligned}\frac{1}{2}m\big(\{y\}\big)\int_J |s(x)-s(y)|\,\rho(dx) + \int_J\int_{s(y)}^{s(x)} m\big((y,s^{-1}(u)\big)\,du\,\rho(dx)\\ \leq T - H^s\left(s(y), \int_J s(x)\,\rho(dx)\right),\end{aligned}$$

where

$$H^s(x,w) = \begin{cases} (x-w)\left(m\big((s^{-1}(x),r)\big) + \frac{1}{2}m\big(\{s^{-1}(x)\}\big)\right), & w < x, \\ 0, & w = x, \\ (w-x)\left(m\big((l,s^{-1}(x))\big) + \frac{1}{2}m\big(\{s^{-1}(x)\}\big)\right), & w > x. \end{cases}$$

Let $\mathcal{A}_3^s(T,y)$ be the measures in $\mathcal{A}^s(T,y)$ which can be written as a weighted sum of at most 3 Dirac measures.

Let $y \in (l,r)$. Then the mapping $\mu \mapsto \nu := \mu \circ s^{-1}$ is a bijection from $\mathcal{A}^s(T,y)$ to $\mathcal{A}^Z\big(T,s(y)\big)$, because

$$\begin{aligned} \int_{J^Z} |x|\,\nu(dx) &= \int_J |s(x)|\mu(dx), \\ \int_{J^Z} x\,\nu(dx) &= \int_J s(x)\mu(dx), \\ \int_{J^Z} q^Z_{s(y)}(x)\,\nu(dx) &= \int_J q^Z_{s(y)}\big(s(x)\big)\,\mu(dx) \\ &= \frac{1}{2}m\big(\{y\}\big)\int_J |s(x)-s(y)|\,\mu(dx) + \int_J\int_{s(y)}^{s(x)} m\big((y,s^{-1}(u)\big)\,du\,\mu(dx). \end{aligned}$$

Furthermore, the number of mass points of μ and ν coincide. Thus, we have proven the following theorem.

Theorem III.7.1. *For a diffusion Y with scale function s we have*

$$V(T,y) = \sup_{\nu\in\mathcal{A}_3^Z(T,s(y))} \int_{\mathbb{R}} f\big(s^{-1}(x)\big)\,\nu(dx) = \sup_{\mu\in\mathcal{A}_3^s(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

III.8 The Lagrangian Approach

The aim of this section is to briefly discuss the Lagrangian approach to constrained optimal stopping problems (cf. Section II.6) in view of Theorem III.1.5 and Theorem III.2.6.

Recall the dual problem

$$w_\lambda(y) = \sup_{\tau\in\mathcal{S}(y)} \mathbb{E}^y[f(Y_\tau) - \lambda\tau], \tag{III.8.1}$$

where $\lambda \in [0,\infty)$, $y \in (l,r)$ and $\mathcal{S}(y) = \bigcup_{T\in[0,\infty)} \mathcal{S}(T,y)$. Since the costs are additive and Y is a Markov process, the continuation and stopping region for the dual problem do not depend on time. Moreover, if f is upper semi-continuous and w is lower semi-continuous, then the first hitting time τ_λ of the stopping region $D_\lambda = \{x \in J \colon w_\lambda(x) = f(x)\}$ is optimal in (III.8.1) provided that $\tau_\lambda < \infty$, $\mathbb{P}^y$-a.s. (see Corollary 2.9 in [55]). Consequently, for every $\lambda \in [0,\infty)$ there exists an optimal stopping time for $w_\lambda(y)$ that can be described as the first exit time of the continuation region. Recall from Section II.6 that if we can identify $\lambda^* \in [0,\infty)$ such that $\mathbb{E}^y[\tau_{\lambda^*}] = T$, then τ_{λ^*} is optimal in the constrained problem (III.1.3) and $V(T,y) = w_{\lambda^*}(y) + \lambda^* T$. In particular, the law of $Y_{\tau_{\lambda^*}}$ under $\mathbb{P}^y$ is a weighted sum of at most two Dirac measures. If a reduction to $\mathcal{A}_2(T,y)$ in (III.1.3) is not possible, then for every $\lambda \in [0,\infty)$ the stopping time τ_λ does not satisfy $\mathbb{E}^y\big[Y_{\tau_\lambda}\big] = T$.

Nevertheless, since the optimal stopping time in the constrained problem (III.1.3) is often an exit time of an interval, we now focus on describing the stopping region D_λ in more detail. Assume that the function f satisfies the conditions of Theorem III.6.2 for every $y \in (l,r)$, i.e. f is upper semi-continuous with $\limsup_{x\uparrow r} \frac{f(x)}{q_y(x)} \le 0$ if $r \notin J$ and $\limsup_{x\downarrow l} \frac{f(x)}{q_y(x)} \le 0$ if $l \notin J$. In particular, there exists an optimal stopping time $\tau \in \mathcal{S}_3(T,y)$ for $V(T,y)$ by Theorem III.6.2 and hence,

$V(T,y)<\infty$ for all $T\in[0,\infty)$ and $y\in(l,r)$. If $\big(q_y(Y_{t\wedge\tau})-(t\wedge\tau_{l,r}\wedge\tau)\big)_{t\in[0,\infty]}$ is a true martingale for every $\tau\in\mathcal{S}(y)$, then the stopping problem (III.8.1) can be rewritten as

$$w_\lambda(y)=\sup_{\tau\in\mathcal{S}(y)}\mathbb{E}^y[f(Y_\tau)-\lambda q_y(Y_\tau)].$$

Fix $z\in(l,r)$ and define $\frac{\partial^0 q_y}{\partial x}(x):=\frac{1}{2}\left(\frac{\partial^+ q_y}{\partial x}(x)+\frac{\partial^- q_y}{\partial x}(x)\right)$. Then (III.2.2) implies that for every $y\in(l,r)$ it holds that

$$\begin{aligned}
w_\lambda(y)&=\sup_{\tau\in\mathcal{S}(y)}\mathbb{E}^y[f(Y_\tau)-\lambda q_y(Y_\tau)]\\
&=\sup_{\tau\in\mathcal{S}(y)}\mathbb{E}^y\left[f(Y_\tau)-\lambda q_z(Y_\tau)-\lambda\frac{\partial^0 q_y}{\partial x}(z)Y_\tau\right]+\lambda\frac{\partial^0 q_y}{\partial x}(z)\,z-\lambda q_y(z)\\
&=\sup_{\tau\in\mathcal{S}(y)}\mathbb{E}^y\left[f_\lambda(Y_\tau)-\lambda\frac{\partial^0 q_y}{\partial x}(z)Y_\tau\right]+\lambda\frac{\partial^0 q_y}{\partial x}(z)\,z-\lambda q_y(z)\\
&=\tilde{w}_{\lambda,y}(y)+\lambda\frac{\partial^0 q_y}{\partial x}(z)\,z-\lambda q_y(z),
\end{aligned}$$

where $f_\lambda(p)=f(p)-\lambda q_z(p)$, $p\in J$, and $\tilde{w}_{\lambda,y}(x):=\sup_{\tau\in\mathcal{S}(x)}\mathbb{E}^x\left[f_\lambda(Y_\tau)-\lambda\frac{\partial^0 q_y}{\partial x}(z)Y_\tau\right]$, $x\in(l,r)$. For a diffusion in natural scale whose accessible boundary points are absorbing, Theorem 3.2 and Remark 3.4 in [53] characterize the value function of an optimal stopping problem with infinite time horizon as the smallest concave function dominating the payoff function on J. Observe that $J\ni p\mapsto f_\lambda(p)-\lambda\frac{\partial^0 q_y}{\partial x}(z)p$ is upper semi-continuous. Moreover, for $\lambda\in(0,\infty)$ the function f_λ is bounded from above and $\limsup_{p\to w}\left\{f_\lambda(p)-\lambda\frac{\partial^0 q_y}{\partial x}(z)p\right\}=-\infty$ for $w\in J\backslash(l,r)$. Hence, the smallest concave function dominating $f_\lambda(p)-\lambda\frac{\partial^0 q_y}{\partial x}(z)p$ is well defined. Moreover, Theorem 3.2 and Remark 3.4 in [53] yield that

$$\begin{aligned}
w_\lambda(y)&=\tilde{w}_{\lambda,y}(y)+\lambda\frac{\partial^0 q_y}{\partial x}(z)\,z-\lambda q_y(z)\\
&=\left(f_\lambda-\lambda\frac{\partial^0 q_y}{\partial x}(z)\mathrm{Id}\right)^{**}(y)+\lambda\frac{\partial^0 q_y}{\partial x}(z)\,z-\lambda q_y(z)\\
&=f_\lambda^{**}(y)+\lambda\frac{\partial^0 q_y}{\partial x}(z)(z-y)-\lambda q_y(z)\\
&=f_\lambda^{**}(y)+\lambda q_z(y),
\end{aligned}$$

where $\mathrm{Id}\colon J\to\mathbb{R}$, $p\mapsto p$, and $g^{**}=(g^*)^*$ denotes the concave biconjugate of a function $g\colon\mathrm{dom}(g)\subseteq\mathbb{R}\to\mathbb{R}$, i.e. $g^*(x)=\inf_{p\in\mathrm{dom}(g)}\{xp-g(p)\}$, $x\in\mathbb{R}$. Note that g^{**} is the smallest concave function dominating g on J. The last equality follows from (III.2.2). Observe that for every $\lambda\in(0,\infty)$ we have that $(l,r)\ni x\mapsto\tilde{w}_{\lambda,y}(x)$ is concave and continuous. Moreover, the payoff function $f_\lambda-\lambda\frac{\partial^0 q_y}{\partial x}(z)\mathrm{Id}$ is upper semi-continuous. The stopping region D_λ for $\tilde{w}_{\lambda,y}$ is given by

$$D_\lambda=\left\{x\in\mathbb{R}\colon f_\lambda(x)-\lambda\frac{\partial^0 q_y}{\partial x}(z)x=\left(f_\lambda-\lambda\frac{\partial^0 q_y}{\partial x}(z)\mathrm{Id}\right)^{**}(x)\right\}=\{x\in\mathbb{R}\colon f_\lambda(x)=f_\lambda^{**}(x)\}.$$

If $\tau_\lambda:=\inf\{t\in[0,\infty)\colon Y_t\in D_\lambda\}$ satisfies $\tau_\lambda<\infty$, $\mathbb{P}^x$-a.s., then a (canonical) optimal stopping time for $\tilde{w}_{\lambda,y}(x)$ is given by τ_λ, see Corollary 2.9 in [55]. Observe that the stopping region D_λ and thus, the stopping time τ_λ are independent of y. Since $\big(g(p)+ap+b\big)^{**}=g^{**}(p)+ap+b$, $a,b\in\mathbb{R}$, it holds that D_λ and τ_λ are also independent of the choice of z. Therefore, τ_λ is optimal for $w_\lambda(y)$ and $w_\lambda(y)$ is independent of the choice of z.

For $y \in (l, r)$ define

$$\begin{aligned} \Phi_y \colon [0,\infty) &\to [0,\infty], \\ \lambda &\mapsto \mathbb{E}^y[\tau_\lambda]. \end{aligned}$$

If $T \in \Phi_y\big([0,\infty)\big)$, then it holds that $\mathbb{E}^y[\tau_{\lambda_*}] = T$ for some $\lambda_* \in [0,\infty)$ and hence, τ_{λ_*} is optimal for $V(T,y)$ with $V(T,y) = w_{\lambda_*}(y) + \lambda_* T$ (cf. Section II.6). In addition, the value of the measure optimization problem (III.2.5) is attained in the set $\mathcal{A}_2(T,y)$.

In the following example we show that a reduction to $\mathcal{A}_2(T,y)$ in (III.1.3) is possible and identify $\lambda_* \in [0,\infty)$ such that τ_{λ_*} is optimal for $V(T,y)$. Afterwards we focus on an example where Φ_y is not surjective and we cannot restrict the set of probability measures in (III.2.5) to $\mathcal{A}_2(T,y)$ for some $(T,y) \in [0,\infty) \times \mathbb{R}$.

Example III.8.1 (cf. Example II.4.6)**.** Let $Y = W$ be a Brownian motion and let $f(x) = |x|$. Then $q_y(x) = (x-y)^2$ and $\big(W_{t\wedge\tau}^2 - (t\wedge\tau)\big)_{t\in[0,\infty]}$ is a martingale for every $\tau \in \mathcal{S}(y)$ by Remark II.4.8. Furthermore, f satisfies the assumptions of Theorem III.6.2. For $z = 0$ it holds that

$$w_\lambda(y) = \sup_{\tau\in\mathcal{S}(y)} \mathbb{E}^y[f(W_\tau) - \lambda\tau] = f_\lambda^{**}(y) + \lambda y^2,$$

where $f_\lambda(x) = |x| - \lambda x^2$. We have for $\lambda > 0$

$$f_\lambda^*(p) = -\frac{(1-|p|)^2}{4\lambda}, \qquad f_\lambda^{**}(z) = \begin{cases} \frac{1}{4\lambda}, & |z| \le \frac{1}{2\lambda}, \\ |z| - \lambda z^2, & |z| \ge \frac{1}{2\lambda}. \end{cases}$$

Therefore, $D_\lambda = \big(-\infty, -\frac{1}{2\lambda}\big] \cup \big[\frac{1}{2\lambda}, \infty\big)$ and the value function w_λ is given by

$$w_\lambda(y) = f_\lambda^{**}(y) + \lambda y^2 = \begin{cases} \frac{1}{4\lambda} + \lambda y^2, & |y| \le \frac{1}{2\lambda}, \\ |y|, & |y| \ge \frac{1}{2\lambda}, \end{cases}$$

with optimal stopping time

$$\tau_\lambda = \tau\left(-\frac{1}{2\lambda}, \frac{1}{2\lambda}\right),$$

where $\tau(a,b) = \inf\{t \in [0,\infty) \colon W_t \notin (a,b)\}$, $a < b$. Thus, $\mathbb{E}^y[\tau_\lambda] = \big(\frac{1}{4\lambda^2} - y^2\big)^+$, which implies that Φ_y is continuous on $(0,\infty)$ with $\Phi_y\big((0,\infty)\big) = [0,\infty)$ if $y \neq 0$ and $\Phi_0\big((0,\infty)\big) = (0,\infty)$. In particular, for $T \in (0,\infty)$ and $y \in \mathbb{R}$ let $\lambda_* = \frac{1}{2\sqrt{T+y^2}}$. Then we have $\mathbb{E}^y[\tau_{\lambda_*}] = T$ and hence,

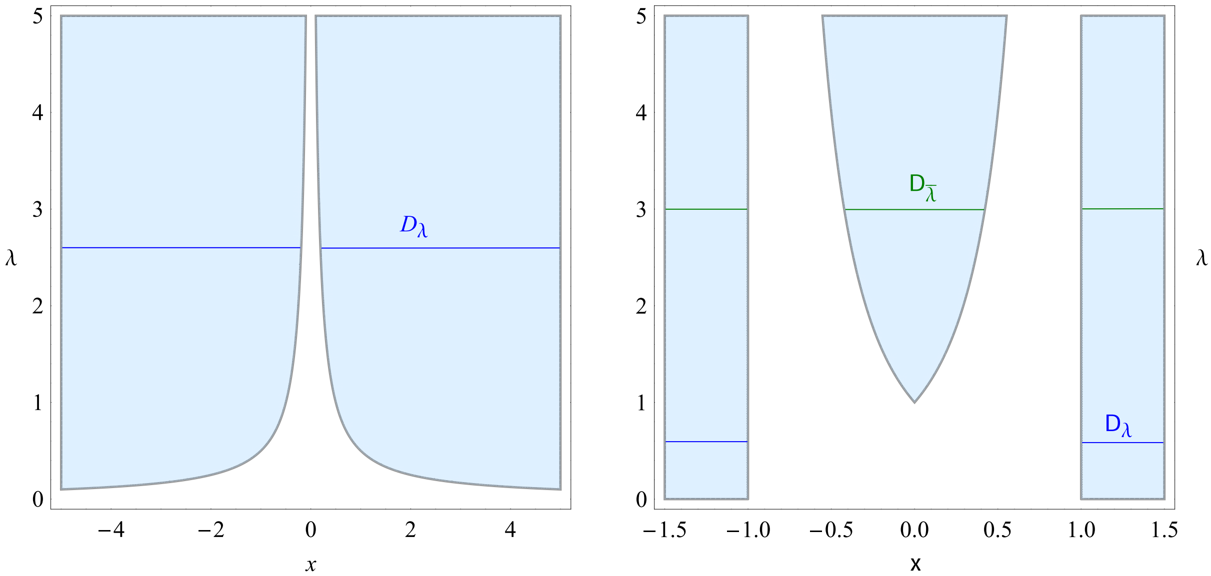
$$V(T,y) = w_{\lambda_*}(y) + \lambda_* T = \sqrt{T+y^2} = \sup_{\mu\in\mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

Observe that for $\lambda = 0$ it holds that $f_0^{**}(x) = \infty$ for all $x \in \mathbb{R}$, hence, $D_0 = \emptyset$ and $\tau_0 = \infty$, $\mathbb{P}^y$-a.s.

Now we provide an example, where there does not exist $\lambda \in [0,\infty)$ with $\mathbb{E}^y[\tau_\lambda] = T$ for some $T \in [0,\infty)$ and $y \in \mathbb{R}$. Moreover, if $T \notin \Phi_y\big([0,\infty)\big)$, then a reduction to weighted sums of at most two Dirac measures in (III.2.5) is possible for some of these $(T,y) \in [0,\infty) \times \mathbb{R}$, but not for all of them.

Example III.8.2 (cf. Example III.4.5)**.** Let $Y = W$ be a Brownian motion starting in $y \in \mathbb{R}$ under $\mathbb{P}^y$ and let $f(x) = \mathbb{1}_{\{|x|\ge 1\}}$. Then similar arguments as in Example III.4.5 show that the value function V of the optimal stopping problem (III.1.3) is given by

$$V(T,y) = \begin{cases} \frac{T}{T+(1-|y|)^2}, & T \le |y|(1-|y|), \\ (T+y^2) \wedge 1, & T \ge |y|(1-|y|). \end{cases}$$



(a) $f(x) = |x|$, see Example III.8.1. (b) $f(x) = \mathbb{1}_{\{|x|\geq 1\}}$, see Example III.8.2.

Figure III.5: The stopping region D_λ for different payoff functions f.

Furthermore, we have that

$$\sup_{\mu\in\mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \begin{cases} \frac{T}{T+(1-|y|)^2}, & T < 1-y^2, \\ (T+y^2)\wedge 1, & T \geq 1-y^2. \end{cases}$$

For $y = 0$ we already know from Example III.4.5 that a reduction to $\mathcal{A}_2(T,0)$ is not possible if $T < 1$. Similarly, one can show that for $|y| < 1$ and $T \in \big(|y|(1-|y|), 1-y^2\big)$ the measure

$$\mu^* = \frac{1}{2}(T+y^2-y)\delta_{-1} + \big(1-T-y^2\big)\delta_0 + \frac{1}{2}(T+y^2+y)\delta_1$$

is optimal in (III.2.5) and that

$$V(T,y) = T+y^2 = \int_{\mathbb{R}} f(x)\mu^*(dx) > \sup_{\mu\in\mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \frac{T}{T+(1-|y|)^2}. \tag{III.8.2}$$

Observe that f fulfills the assumptions of Theorem III.6.2. For $\lambda \in [0,\infty)$ let $f_\lambda(x) = f(x) - q_0(x) = f(x) - \lambda x^2$ (see Example III.8.1). We obtain for $\lambda = 0$

$$f_0^*(p) = \begin{cases} -1, & p = 0, \\ -\infty, & p \neq 0, \end{cases} \qquad\qquad f_0^{**}(z) = 1,$$

and for $\lambda \in (0,1)$

$$f_\lambda^*(p) = \begin{cases} -\frac{p^2}{4\lambda} - 1, & |p| \geq 2\lambda, \\ \lambda - |p| - 1, & |p| \leq 2\lambda, \end{cases} \qquad\qquad f_\lambda^{**}(z) = \begin{cases} 1-\lambda, & |z| \leq 1, \\ 1-\lambda z^2, & |z| \geq 1. \end{cases}$$

For $\lambda \geq 1$ it holds that

$$f_\lambda^*(p) = \begin{cases} -\frac{p^2}{4\lambda} - 1, & |p| \geq 2\lambda, \\ \lambda - |p| - 1, & 2(\lambda - \sqrt{\lambda}) \leq |p| \leq 2\lambda, \\ -\frac{p^2}{4\lambda}, & 0 \leq |p| \leq 2(\lambda - \sqrt{\lambda}), \end{cases}$$

$$f_\lambda^{**}(z) = \begin{cases} -\lambda z^2, & |z| \leq 1 - \frac{1}{\sqrt{\lambda}}, \\ -2(\lambda - \sqrt{\lambda})|z| + \left(1 - \sqrt{\lambda}\right)^2, & 1 - \frac{1}{\sqrt{\lambda}} \leq |z| \leq 1, \\ 1 - \lambda z^2, & |z| \geq 1. \end{cases}$$

Therefore, the stopping region is given by

$$D_\lambda = (-\infty, -1] \cup [1, \infty), \qquad \lambda \in [0, 1),$$

$$D_\lambda = (-\infty, -1] \cup \left[-1 + \frac{1}{\sqrt{\lambda}}, 1 - \frac{1}{\sqrt{\lambda}}\right] \cup [1, \infty), \qquad \lambda \geq 1.$$

The value function $w_\lambda(y) = f_\lambda^{**}(y) + \lambda y^2$ is given by

$$w_\lambda(y) = \begin{cases} 1 - \lambda + \lambda y^2, & |y| \leq 1, \\ 1, & |y| \geq 1, \end{cases} \qquad \lambda \in [0, 1),$$

$$w_\lambda(y) = \begin{cases} 0, & |y| \leq 1 - \frac{1}{\sqrt{\lambda}}, \\ \left[\sqrt{\lambda}(|y| - 1) + 1\right]^2, & 1 - \frac{1}{\sqrt{\lambda}} \leq |y| \leq 1, \\ 1, & |y| \geq 1, \end{cases} \qquad \lambda \geq 1,$$

with optimal stopping times

$$\tau_\lambda^* = \tau(-1, 1), \qquad \lambda \in [0, 1),$$

$$\tau_\lambda^* = \max\left\{\tau\left(-1, -1 + \frac{1}{\sqrt{\lambda}}\right), \tau\left(1 - \frac{1}{\sqrt{\lambda}}, 1\right)\right\} \qquad \lambda \geq 1.$$

We conclude that $\mathbb{E}^y\left[\tau_\lambda^*\right] = 0$ for all $|y| \geq 1$. For $|y| < 1$ we have

$$\mathbb{E}^y\left[\tau_\lambda^*\right] = \begin{cases} 1 - y^2, & \lambda \in [0, 1), \\ (1 - |y|)\left(\frac{1}{\sqrt{\lambda}} - 1 + |y|\right), & 1 \leq \lambda < \frac{1}{(1-|y|)^2}, \\ 0, & \lambda \geq \frac{1}{(1-|y|)^2}. \end{cases}$$

Thus,

$$\Phi_y\big([0, \infty)\big) = \begin{cases} \{0\}, & |y| \geq 1, \\ \big[0, |y|(1 - |y|)\big] \cup \big\{1 - y^2\big\}, & |y| < 1. \end{cases}$$

For $|y| \geq 1$ stopping directly yields a payoff of 1 and thus, is optimal, because f is bounded above by 1.

Now let $|y| < 1$ and $T > 1 - y^2$. Then the method described above does not provide an optimal stopping time for $V(T, y)$. But f is constant and maximal on $\mathbb{R}\backslash(-1, 1)$ and the expected time to reach either -1 or 1 is smaller or equal to T. Therefore, the stopping times $\tau(a, b)$ with $a \leq -1$, $b \geq 1$ and $(b - y)(y - a) \leq T$ are optimal for $V(T, y)$.

Let $|y| < 1$ and $T \in \big(|y|(1-|y|), 1-y^2\big)$. Then we cannot recover V from w_λ by using the stopping times τ^*_λ. Note that by (III.8.2) a reduction to $\mathcal{A}_2(T,y)$ is not possible. Nevertheless, there exists an optimal stopping time τ_* for $w_1(y)$ with $\mathbb{E}^y[\tau_*] = T$. Indeed, let

$$\tau_* = \tau(-1,b) + \mathbb{1}_{\{W_{\tau(-1,b)}=b\}} \inf\big\{t \in [0,\infty) \colon W_{\tau(-1,b)+t} \notin (0,1)\big\},$$

where $b = \frac{T+y^2+y}{2-T-y^2+y}$. Then τ_* satisfies $\mathbb{E}^y[\tau_*] = T$ and $\mathbb{E}^y[f_1(W_{\tau_*})] + y^2 = y^2 = w_1(y)$ and thus τ_* is optimal for $V(T,y)$. Notice that τ_* is an embedding in W under $\mathbb{P}^y$ of the optimal measure

$$\mu^* = \frac{1}{2}\big(T+y^2-y\big)\delta_{-1} + \big(1-T-y^2\big)\delta_0 + \frac{1}{2}\big(T+y^2+y\big)\delta_1$$

for $V(T,y)$, see the proof of Theorem III.1.5.

In general, it can be challenging to determine λ^* and an optimal stopping time τ_{λ^*} for $w_{\lambda^*}(y)$ with $\mathbb{E}^y[\tau_{\lambda^*}] = T$.

IV. Sequential Testing – Optimal Exit Strategies

This chapter serves as an extensive example for the results of Chapter II and III.

When deciding between two simple hypotheses, e.g. testing clinical effectiveness of a medicament, an economic agent collects information. The incoming information often does not provide a significant result or a decision rule within a sharp time horizon T. Then the question arises whether to continue the observation process after time T or to abandon collecting information with no result. The agent has to weigh additional observation costs against the loss of benefits associated to a significant result. In particular, if the observations until time T suggest one outcome, but are not significant enough, then the decision maker intends to continue collecting information. Conversely, if the first observations are significant, it is not necessary to exploit the full time horizon T. Thus, we introduce an average time constraint until the agent has to terminate the observation process.

The agent aims at maximizing the payoff associated to the two hypotheses. Moreover, we assume that if she quits collecting information without a decision, then she gains nothing. The aggregated incoming information is modeled by a random walk, i.e. all positive increments account for hypothesis 1 (e.g. the medicine is effective) whereas the negative increments contribute for the second hypothesis (e.g. the medicine has no verifiable effect). Thus, the agent wants to detect whether the random walk has a positive or negative drift. For this purpose we consider the continuous time sequential testing model from Chapter VI.21 in [55]. In the sequential testing model the belief about the probabilities of each hypothesis, the so-called a posteriori probabilities, are updated according to the observations exactly at the time they are available.

The average time constraint entails that the agent stops the observation process at a finite point in time which depends on the scenario. In addition, we show that the agent can restrict to stopping times such that the law of the a posteriori probability process at the stopping time is a weighted sum of at most 3 Dirac measures. This implies that allowing for three possible outcomes, i.e. accepting hypothesis 1 or 2 or quitting the decision making process with *no* result, increases the expected payoff of the agent. If in our setting stopping at 3 points yields a higher payoff than stopping at 2 points, the optimal 3 points do not depend on the average time constraint nor on the a priori probabilities of the two hypotheses. The *same* 3 points suffice. On the contrary, if stopping at two points is optimal, then exactly one of the mass points in the optimal measure is independent of the time constraint and the a priori probabilities.

IV.1 The Sequential Testing Model

We now describe the sequential testing model from Chapter VI.21 in [55]. Let $\big(\Omega, \mathcal{F}, (\mathbb{P}^y)_{y\in[0,1]}\big)$ be a probability-statistical space. In the Bayesian formulation the probability measures $\mathbb{P}^y$, $y \in [0, 1]$, are given by

$$\mathbb{P}^y = y\mathbb{P}^1 + (1 - y)\mathbb{P}^0.$$

Let $W = (W_t)_{t\in[0,\infty)}$ be a Brownian motion starting in 0 under every $\mathbb{P}^y$. Furthermore, let θ be a random variable independent of W under every $\mathbb{P}^y$ with $\mathbb{P}^y[\theta = 1] = y$ and $\mathbb{P}^y[\theta = 0] = 1 - y$. Let

$$X_t = \theta\kappa t + \sigma W_t, \quad t \in [0,\infty),$$

where $\kappa \in \mathbb{R}\backslash\{0\}$ and $\sigma^2 > 0$. Then $\mathbb{P}^y[X \in \cdot \,|\, \theta = i] = \mathbb{P}^i[X \in \cdot]$, $i \in \{0,1\}$, is the law of a Brownian motion with drift $i\kappa$ and diffusion coefficient σ. For $t \in [0,\infty)$ let $\mathcal{F}_t^X = \sigma(X_s \colon 0 \le s \le t)$ be the σ-algebra generated by $(X_s)_{s\in[0,t]}$. One aims at deriving the value of θ from continuously observing the process X. Note that the a priori probabilities of the statistical hypotheses

$$H_0 : \theta = 0 \quad \text{and} \quad H_1 : \theta = 1$$

are given by $1 - y$ and y under $\mathbb{P}^y$. Define the a posteriori probability process $(Y_t)_{t\in[0,\infty)}$ by

$$Y_t := \mathbb{P}^y[\theta = 1 \,|\, \mathcal{F}_t^X].$$

According to Theorem 7.1 in [41] the likelihood ratio process $(\varphi_t)_{t\in[0,\infty)}$ defined as the Radon-Nikodým derivative of the measure $\mathbb{P}^1$ with respect to $\mathbb{P}^0$ on $\mathcal{F}_t^X$ satisfies

$$\varphi_t := \frac{d\mathbb{P}^1}{d\mathbb{P}^0}\bigg|_{\mathcal{F}_t^X} = \exp\left(\frac{\kappa}{\sigma^2}\left(X_t - \frac{\kappa}{2}t\right)\right).$$

Moreover, we conclude from [63, p. 181] that

$$Y_t = \left(\frac{y}{1-y}\varphi_t\right) \Big/ \left(1 + \frac{y}{1-y}\varphi_t\right)$$

and that $(Y_t)_{t\in[0,\infty)}$ solves

$$dY_t = \frac{\kappa}{\sigma} Y_t(1 - Y_t)\, d\widetilde{W}_t, \quad Y_0 = y, \tag{IV.1.1}$$

where

$$\widetilde{W}_t = \frac{1}{\sigma}\left(X_t - \kappa\int_0^t Y_s ds\right)$$

is a standard Brownian motion with respect to $(\mathcal{F}_t^X)$ and $\mathbb{P}^y$, see Theorem 7.12 and Theorem 9.1 in [41]. Moreover, the filtration $(\mathcal{F}_t^Y)_{t\in[0,\infty)}$ generated by $(Y_t)_{t\in[0,\infty)}$ coincides with $(\mathcal{F}_t^X)_{t\in[0,\infty)}$. Observe that $(Y_t)_{t\in[0,\infty)}$ is a regular continuous strong Markov process under $\mathbb{P}^y$ with state space $J = (0,1)$. From (IV.1.1) we conclude that if Y_t is close to 0 or 1, then it is very unlikely that the a posteriori probability changes much in small time intervals.

We introduce a threshold $\alpha \in \left(0, \frac{1}{2}\right)$ and assume that the observations are significant enough to accept one of the hypotheses if the a posteriori process exceeds $1-\alpha$ or falls below α. In particular, we impose that the decision maker can accept H_0 and H_1 at time $t \in [0,\infty)$ only if $Y_t \le \alpha$ and $Y_t \ge 1 - \alpha$, respectively. Detecting a drift 0 or κ can be of different value for the decision maker. We normalize the gain for accepting H_0 and assume that H_1 yields a payoff $\beta \ge 1$. On the other hand, if the decision maker cannot accept H_0 nor H_1 at the time she terminates the observation, then she gains nothing. Thus, the payoff associated to the two possible drift rates provides an incentive to accept H_0 and H_1; and the agent prefers to accept H_1.

For $\beta \ge 1$ let the payoff function for the agent be given by

$$f(x) = \mathbb{1}_{(0,\alpha]}(x) + \beta\mathbb{1}_{[1-\alpha,1)}(x).$$

We presume that the agent faces the optimal stopping problem from Chapter III: She wants to find a stopping time τ with $\mathbb{E}[\tau] \le T$ such that the expected payoff depending on the a posteriori probability process is maximized. The value function $V \colon [0,\infty) \times (0,1) \to \mathbb{R}$ is given by

$$V(T, y) = \sup\left\{\mathbb{E}^y[f(Y_\tau)] \colon \tau \ (\mathcal{F}_t^Y)\text{-stopping time},\ \mathbb{E}^y[\tau] \le T\right\}. \tag{IV.1.2}$$

The payoff function f does not respect whether the observations are more significant than the given threshold α. Nevertheless, the agent does not have to accept H_0 or H_1 at time t if $Y_t \leq \alpha$ or $Y_t \geq 1-\alpha$, respectively. She can also continue observing the process X and hope for significant results for the other hypothesis.

It may happen that the a posteriori process stays in $(\alpha, 1-\alpha)$ for a long time and thus, the agent cannot accept H_0 or H_1 within a given sharp time horizon. To avoid waiting arbitrarily long for a significant result, we impose a constraint on the average time until the observation process has to be terminated. The decision maker observes the process X until a stopping time τ and at τ she either accepts H_0 and H_1 if $Y_\tau \leq \alpha$ and $Y_\tau \geq 1-\alpha$, respectively, or she quits collecting information with no result for the value of θ. In particular, if she accepts one of the hypotheses at the stopping time τ, then the drift is assumed to equal 0 or κ.

For the stopping problem (IV.1.2) it is sufficient to consider only stopping times such that the law of the process at the stopping time is a weighted sum of at most 3 Dirac measures, cf. Theorem III.1.5. In general, a reduction to 2 mass points is not possible as we have seen in Example III.4.5. Hence, the optimal stopping time, which exists by Theorem III.6.2, is a consecutive exit time by Theorem III.1.5. Recall that in optimal stopping problems with finite or infinite time horizon one can reduce the set of stopping times to first exit times of intervals, see e.g. Corollary 2.9 in [55]. In particular, the process at the stopping time has at most 2 mass points. Therefore, compared to the unconstrained problems, the expectation constraint leads to a systematical difference in the optimal stopping time. Nevertheless, an optimal stopping time in the constrained problem is often an exit time of an interval. Thus, we introduce an auxiliary optimal stopping problem, in which we only allow for exit times $\tau = \inf\{t \in [0,\infty) \colon Y_t \notin (a,b)\}$, $0 < a \leq b < 1$, satisfying the expectation constraint. In particular, the law of Y_τ has at most two mass points. For $(T,y) \in [0,\infty) \times (0,1)$ let

$$V_2(T,y) = \sup\left\{\mathbb{E}^y[f(Y_\tau)] \colon \mathbb{E}^y[\tau] \leq T,\ \mathrm{Law}(Y_\tau) \text{ weighted sum of at most 2 Dirac measures}\right\}. \tag{IV.1.3}$$

The value function $V_2(T,y)$ and an optimal stopping time for (IV.1.3) help to derive $V(T,y)$.

In Section IV.3 and Section IV.4 we show that if the value of the optimal stopping problem increases if we allow not only for exit times of intervals but for consecutive exit times, i.e. $V(T,y) > V_2(T,y)$, then there exists an optimal stopping time τ^* for (IV.1.2) such that the law of Y_{τ^*} under $\mathbb{P}^y$ is a weighted sum of Dirac measures in $\alpha, b^*, 1-\alpha$, where $b^* \in \left(\alpha, \frac{1}{2}\right]$. The mass points are independent of T and y. Hence, the same three points suffice in (IV.1.2).

To apply the results from Chapter III, we first derive the function q_y, $y \in (0,1)$, that is associated to the process $(Y_t)_{t\in[0,\infty)}$. Remark III.1.2 implies that

$$q_y(x) = \begin{cases} \dfrac{2\sigma^2}{\kappa^2}\left((2x-1)\log\left(\dfrac{(1-y)x}{(1-x)y}\right) + \dfrac{1-2y}{y(1-y)}(x-y)\right), & x \in (0,1), \\ \infty, & x \notin (0,1). \end{cases}$$

If we impose that both hypotheses have probability $\frac{1}{2}$ at the beginning of the observation, i.e. $y = \frac{1}{2}$, then the function $q_{\frac{1}{2}}$ simplifies to

$$q(x) := q_{\frac{1}{2}}(x) = \begin{cases} \dfrac{2\sigma^2}{\kappa^2}(2x-1)\log\left(\dfrac{x}{1-x}\right), & x \in (0,1), \\ \infty, & x \notin (0,1). \end{cases}$$

In particular, we have $q(x) = q(1-x)$, $x \in \mathbb{R}$. Recall from (III.2.2) that for every $y \in (0,1)$ it holds that

$$q_y(x) = q(x) - q(y) - (x-y)q'(y), \qquad x \in (0,1). \tag{IV.1.4}$$

In the following we first state the value function of the primal and the auxiliary optimal stopping problem for $y = \frac{1}{2}$ and examine the dependence of the value function and the optimal stopping times on the exogenous parameters α, β, κ and σ in Section IV.2. In Section IV.3 we collect the value functions V_2 and V and the optimal stopping times for a general a priori probability of $\{\theta = 1\}$ and prove these results in Section IV.4.

IV.2 Optimal Stopping Rules when starting without Bias

In this section we assume that the agent has no bias at the beginning of the observation process, i.e. $y = \frac{1}{2}$. First we state the value function and an optimal stopping time for the case where hypothesis 1 has a higher payoff. Then we focus on the case where both hypotheses yield the same payoff, i.e. $\beta = 1$.

Since the a priori probability of $\{\theta = 1\}$ is fixed, we write in the following $V(T)$ and $V_2(T)$ instead of $V\left(T, \frac{1}{2}\right)$ and $V_2\left(T, \frac{1}{2}\right)$, respectively.

IV.2.1 H_1 yields a higher Payoff

Recall that the payoff of hypothesis H_1 is given by $\beta \geq 1$. Moreover, before the observation starts the agent believes that the drift 0 and κ both occur with probability $\frac{1}{2}$.

Let $\tau(a,b) = \inf\{t \in [0,\infty) \colon Y_t \notin (a,b)\}$, $0 < a < b < 1$. Observe that the expected time until the process $(Y_t)_{t\in[0,\infty)}$ hits either α or $1-\alpha$ for the first time is given by

$$\mathbb{E}^{\frac{1}{2}}[\tau(\alpha, 1-\alpha)] = \mathbb{E}^{\frac{1}{2}}[q(Y_\tau)] = \frac{\frac{1}{2}-\alpha}{1-2\alpha} q(\alpha) + \frac{\frac{1}{2}-\alpha}{1-2\alpha} q(1-\alpha) = q(\alpha),$$

see Lemma 2.2 in [5]. If the upper bound for the expected time horizon is larger than $q(\alpha)$, then the agent obtains a payoff of $\frac{1}{2}(\beta+1)$ by stopping at α and $1-\alpha$. To increase the payoff, she can increase the probability to attain $1-\alpha$ by stopping at $1-\alpha$ and a point $a^* \in (0,\alpha]$ such that $\tau(a^*, 1-\alpha)$ satisfies the expectation constraint.

If the time horizon is smaller than $q(\alpha)$, then two cases can occur. If the constraint is not too small, i.e. $T \in \left(T^*, q(\alpha)\right)$ for some $T^* \in \left[0, q(\alpha)\right)$, then stopping at three points yields a higher payoff than stopping at two points. The optimal measure has mass points α, $1-\alpha$ and $b^* \in (\alpha, 1-\alpha)$, where b^* is independent of the time constraint. If $T \leq T^*$, then stopping at two points is optimal. The optimal stopping rule is as follows: If the time constraint is too small to reach b^* and $1-\alpha$ in expectation, then we stop at $1-\alpha$ and a point $a^* = a^*(T) \in \left(b^*, \frac{1}{2}\right)$ such that $\mathbb{E}^{\frac{1}{2}}[\tau(a^*, 1-\alpha)] = T$. In particular, the point a^* decreases to b^* for $T \nearrow T^*$.

For $T \in \left(T^*, q(\alpha)\right)$ it is optimal to use a consecutive exit time: First stop at α and $b_1 \in \left(\frac{1}{2}, 1-\alpha\right)$ and if the process attains b_1 before α, then continue until the process either hits b^* or $1-\alpha$.

In order to state the value function more precisely, let

- $b^* \in \left(\alpha, \frac{1}{2}\right]$ be the unique solution on $[\alpha, 1-\alpha]$ of
$$\ell(b) := (\beta-1)\big(q(\alpha) - q(b)\big) + (1-\alpha-b+b\beta-\alpha\beta)q'(b) = 0,$$
- $T^* = \frac{\frac{1}{2}-b^*}{1-\alpha-b^*} q(\alpha) + \frac{\frac{1}{2}-\alpha}{1-\alpha-b^*} q(b^*)$,
- $a^*(T)$ be the unique solution of $\frac{\frac{1}{2}-\alpha}{1-\alpha-a} q(a) + \frac{\frac{1}{2}-a}{1-\alpha-a} q(\alpha) = T$ on $\left(0, \frac{1}{2}\right]$.

Now we can formulate the main result of this section.

Theorem IV.2.1. *The value function V of the optimal stopping problem* (IV.1.2) *is given by*

$$V(T) = \begin{cases} V_2(T), & T \in [0, T^*] \cup [q(\alpha), \infty), \\ \dfrac{T - q(b^*)}{q(\alpha) - q(b^*)} + (\beta - 1)\left(\dfrac{1}{2} - \dfrac{\big(q(\alpha) - T\big)(b^* - \alpha)}{(1-2\alpha)\big(q(\alpha) - q(b^*)\big)}\right), & T \in \left(T^*, q(\alpha)\right), \end{cases}$$

where

$$V_2(T) = \begin{cases} \beta \dfrac{\frac{1}{2} - a^*(T)}{1-\alpha-a^*(T)}, & T \in \big[0, q(\alpha)\big), \\ 1 + (\beta - 1)\dfrac{\frac{1}{2} - a^*(T)}{1-\alpha-a^*(T)}, & T \geq q(\alpha). \end{cases}$$

For $T \in [0, T^*] \cup [q(\alpha), \infty)$ the stopping time $\tau\big(a^*(T), 1-\alpha\big)$ is optimal for $V(T)$. For $T \in \big(T^*, q(\alpha)\big)$ an optimal stopping time for $V(T)$ is given by

$$\tau^* = \tau(\alpha, b_1) + \mathbb{1}_{\{Y_{\tau(\alpha,b_1)} = b_1\}} \inf\big\{t \in [0,\infty) \colon Y_{\tau(\alpha,b_1)+t} \notin (b^*, 1-\alpha)\big\},$$

where $b_1 \in \big(\frac{1}{2}, 1-\alpha\big)$ is given by

$$b_1 = 1 - \alpha - \frac{(1-2\alpha)(1-\alpha-b^*)\big(q(\alpha) - T\big)}{\big(\frac{3}{2} - 2\alpha - b^*\big)q(\alpha) - \big(\frac{1}{2} - \alpha\big)q(b^*) - (1-\alpha-b^*)T}.$$

Here we do not provide a proof because the statement follows from more general results (Lemma IV.3.1 and Theorem IV.3.2) in Section IV.3.

If the payoff for hypothesis 1 is larger than for hypothesis 0, then the expected payoff depends on three cases for the average time constraint. If the a posteriori probability process can attain, in expectation, α and $1-\alpha$ within the given time horizon, then it is optimal for the agent to wait until she either can accept H_1 for the first time or the a posteriori probability process falls below $a^*(T) \leq \alpha$. Although she can assume that the drift equals 0 if the a posteriori process is less or equal to α, she continues observing the signals. This is due to the fact that she hopes for enough positive signals in the remaining time such that the observations suggest a drift κ. To maximize the payoff the agent always wants the signals to indicate a drift κ and the a posteriori probability process to hit $1-\alpha$. In the case of very small time horizons, if a drift κ seems unlikely given the first signals, then she quits observing the process with no result. Here unlikely means that the a posteriori probability for the drift being κ is less or equal to $a^*(T)$.

Finally, if the time constraint is large enough but Y cannot reach both α and $1-\alpha$ in expectation, then the decision maker stops observing the signals if the a posteriori probability process equals α or $1-\alpha$ for the first time or b^* for the first time after it hits a level $b_1 \in \big(\frac{1}{2}, 1-\alpha\big)$. In particular, this corresponds to stopping at a time τ^* if the agent is convinced that the drift is either 0 or κ and she quits the observation process with no result if the a posteriori probability process first attains a level b_1 and then goes back to b^* before hitting $1-\alpha$, roughly speaking she stops at b^* if it takes too long to start the observation process afresh after it has first attained b_1, then falls below $\frac{1}{2}$ and attains b^*.

IV.2.2 Both Hypotheses have the same Payoff

In this section we assume that both hypotheses yield the same payoff. If $\beta = 1$, then the result of Section IV.2.1 simplifies.

Corollary IV.2.2. *For $\beta = 1$ the value function of the optimal stopping problem* (IV.1.2) *is given by*

$$V(T) = \begin{cases} \dfrac{T}{q(\alpha)}, & T < q(\alpha), \\ 1, & T \geq q(\alpha). \end{cases}$$

Let $T \in \big(0, q(\alpha)\big)$. Then the following consecutive exit time τ^* is optimal in (IV.1.2).

$$\tau^* = \tau(\alpha, b_1) + \mathbb{1}_{\{Y_{H(\alpha,b_1)} = b_1\}} \inf\left\{t \in [0,\infty) \colon Y_{\tau(\alpha,b_1)+t} \notin \left(\frac{1}{2}, 1-\alpha\right)\right\},$$

where $b_1 = \frac{q(\alpha) - \alpha T}{2q(\alpha) - T}$.

Observe that for $\beta = 1$ we only have two cases for the time constraint. If T is not too big, i.e. if T is smaller than the expected time to reach both α and $1-\alpha$, then the maximal expected payoff is not attained by a stopping time such that the process at the stopping time has two mass points. Three points are necessary: Similar to the case $\beta \geq 1$ the agent stops if the a posteriori probability process equals α or $1-\alpha$ for the first time or $\frac{1}{2}$ for the first time after it hits a level $b_1 \in \left(\frac{1}{2}, 1-\alpha\right)$. In particular, she terminates the observation process with no result if the a posteriori probability process first attains b_1 and then goes back to $\frac{1}{2}$ before hitting α. Hence, it takes too long to start the observation process once again.

IV.2.3 Dependence on the Parameters

We examine how the optimal stopping rule and the expressions used to obtain the value function V in Section IV.2.1 change in the parameters and interpret these changes. More precisely, we focus on the dependence of $q, a^*(T), V_2, b^*, T^*$ and V on α, β, κ and σ. We first summarize and comment on the dependence on the parameters and prove them afterwards. We use $\uparrow$ and $\downarrow$ to show that the expression in the left column increases respectively decreases when the parameter increases. The symbol $\circ$ represents that there is no dependence on the parameter.

	α	β	$\left\lvert\frac{\kappa}{\sigma}\right\rvert$
q	$\circ$	$\circ$	$\downarrow$
$a^*(T)$	$\downarrow$	$\circ$	$\downarrow$
$V_2(T)$	$\uparrow$	$\uparrow$	$\uparrow$
b^*	$\uparrow$	$\downarrow$	$\circ$
T^*	$\downarrow$	$\uparrow$	$\downarrow$
$V(T)$	$\uparrow$	$\uparrow$	$\uparrow$

The threshold α allows the agent to assume that the drift is 0 if the a posteriori probability process is less or equal to α and to presume that the drift equals κ if the process is greater or equal to $1-\alpha$. If α increases, then she can already decide earlier on the drift's value. Hence, the payoff function f increases and, thus, the value functions V_2 and V increase. Furthermore, if α increases, then stopping at three points yields a higher payoff than stopping at two points for smaller time horizon T, because of the increasing payoff function. This explains why T^* decreases in α.

If the gain for hypothesis 1 increases, i.e. β increases, then the value functions V_2 and V increase. For higher values of β it is better to assign more mass to the point $1-\alpha$ than putting mass in both α and $1-\alpha$. In particular, the decision maker accepts to quit collecting information without a result and hence, to obtain nothing in order to hit $1-\alpha$ with a higher probability and thus, to increase the expected payoff. Therefore, we have to impose a higher constraint on the average waiting time to allow for three possible outcomes, i.e. T^* increases. Since the expected time until the a posteriori probability process attains α and $1-\alpha$ does not change, the length of the time interval for which three outcomes are optimal decreases in β.

The ratio $\frac{\kappa}{\sigma}$ is a measure for the strength of the observable signals: If κ is large compared to the diffusion coefficient σ of the process X, then the influence of a drift κ will predominate and the agent can conclude the value of θ after a short observation time. On the other hand, if $\frac{\kappa}{\sigma}$ is small, it becomes more difficult for the agent to decide whether she observes a drift or the effect of the noise. Thus, if the strength of the signal increases, the value function increases. Observe that the function q decreases in $\left\lvert\frac{\kappa}{\sigma}\right\rvert$. The third stopping point b^* does not change in $\left\lvert\frac{\kappa}{\sigma}\right\rvert$, because all expressions are scaled with the same factor. Then the expected time to attain the three points α, b^* and $1-\alpha$ is smaller and thus, T^* is decreasing. Since q decreases in $\left\lvert\frac{\kappa}{\sigma}\right\rvert$, when using two consecutive exit times, the auxiliary stopping point can increase and the constraint is still satisfied. In particular, an increasing auxiliary stopping point b_1 implies that the mass in the optimal measure in α and $1-\alpha$ increases and hence, the payoff increases. Furthermore, since b^* does not depend on the strength of the signal, the mass in α and $1-\alpha$ increase in $\left|\frac{\kappa}{\sigma}\right|$.

Dependence on the Threshold α

The function q is independent of α, but note that $\mathbb{E}^{\frac{1}{2}}[\tau(\alpha, 1-\alpha)] = q(\alpha)$ decreases in $\alpha \in \left(0, \frac{1}{2}\right)$ with $\lim_{\alpha\downarrow 0} q(\alpha) = \infty$ and $q\left(\frac{1}{2}\right) = 0$. The payoff function f is increasing in α, thus, also the value functions V_2 and V increase in α.

In the following we write V_2^α and $a^*(T,\alpha)$ to emphasize the dependence of V_2 and $a^*(T)$ on α. First observe that

$$\frac{\frac{1}{2}-\alpha}{1-\alpha-a}q(a) + \frac{\frac{1}{2}-a}{1-\alpha-a}q(\alpha) \tag{IV.2.1}$$

strictly decreases in $a \in \left(0, \frac{1}{2}\right)$ for fixed $\alpha \in \left(0, \frac{1}{2}\right)$ as well as in α for fixed a. Hence, for $0 < \alpha < \gamma < \frac{1}{2}$ and $T \in (0, \infty)$ we have

$$\begin{aligned} T &= \frac{\frac{1}{2}-\alpha}{1-\alpha-a^*(T,\alpha)}q\big(a^*(T,\alpha)\big) + \frac{\frac{1}{2}-a^*(T,\alpha)}{1-\alpha-a^*(T,\alpha)}q(\alpha) \\ &> \frac{\frac{1}{2}-\gamma}{1-\gamma-a^*(T,\alpha)}q\big(a^*(T,\alpha)\big) + \frac{\frac{1}{2}-a^*(T,\alpha)}{1-\gamma-a^*(T,\alpha)}q(\gamma), \end{aligned}$$

which implies that $a^*(T,\gamma) < a^*(T,\alpha)$ by the definition of $a^*(T,\gamma)$. Thus, $a^*(T,\alpha)$ is strictly decreasing in α. Furthermore, it holds that $a^*(T,\alpha) \to 0$ as $\alpha \nearrow \frac{1}{2}$. Indeed, let $\varepsilon \in \left(0, \frac{1}{2}\right)$ and assume that $a^*(T,\alpha) \geq \varepsilon$ for all $\alpha \in \left(0, \frac{1}{2}\right)$. Since (IV.2.1) is decreasing in a, it follows that

$$\begin{aligned} T &= \frac{\frac{1}{2}-\alpha}{1-\alpha-a^*(T,\alpha)}q\big(a^*(T,\alpha)\big) + \frac{\frac{1}{2}-a^*(T,\alpha)}{1-\alpha-a^*(T,\alpha)}q(\alpha) \\ &\leq \frac{\frac{1}{2}-\alpha}{1-\alpha-\varepsilon}q(\varepsilon) + \frac{\frac{1}{2}-\varepsilon}{1-\alpha-\varepsilon}q(\alpha) \\ &\xrightarrow[\alpha\uparrow\frac{1}{2}]{} 0. \end{aligned}$$

Hence, $\lim_{\alpha\uparrow\frac{1}{2}} a^*(T,\alpha) = 0$. Similarly, one shows that $\lim_{\alpha\downarrow 0} a^*(T,\alpha) = \frac{1}{2}$.

Since $\frac{\frac{1}{2}-a}{1-\alpha-a}$ strictly decreases in a, $a^*(T,\alpha)$ is strictly decreasing in α and $\frac{\frac{1}{2}-a}{1-\alpha-a}$ is strictly increasing in α, we deduce for $0 < \alpha < \gamma < \frac{1}{2}$ that

$$\frac{\frac{1}{2}-a^*(T,\alpha)}{1-\alpha-a^*(T,\alpha)} < \frac{\frac{1}{2}-a^*(T,\gamma)}{1-\alpha-a^*(T,\gamma)} < \frac{\frac{1}{2}-a^*(T,\gamma)}{1-\gamma-a^*(T,\gamma)}.$$

Furthermore, it holds that $q(\gamma) < q(\alpha)$. Thus, for $T \in \big(0, q(\gamma)\big)$ we have

$$V_2^\alpha(T) = \beta\frac{\frac{1}{2}-a^*(T,\alpha)}{1-\alpha-a^*(T,\alpha)} < \beta\frac{\frac{1}{2}-a^*(T,\gamma)}{1-\gamma-a^*(T,\gamma)} = V_2^\gamma(T).$$

Similarly, we conclude that $V_2^\alpha(T) < V_2^\gamma(T)$ for $T \in \big[q(\alpha), \infty\big)$. For $T \in \big[q(\gamma), q(\alpha)\big)$ it holds that

$$V_2^\alpha(T) = \beta\frac{\frac{1}{2}-a^*(T,\alpha)}{1-\alpha-a^*(T,\alpha)} < \beta\frac{\frac{1}{2}-a^*(T,\gamma)}{1-\gamma-a^*(T,\gamma)} + \frac{\frac{1}{2}-\gamma}{1-\gamma-a^*(T,\gamma)} = V_2^\gamma(T).$$

To summarize, $V_2^\alpha(T)$ is strictly increasing in α for $T \in (0,\infty)$. In addition, we have that $\lim_{\alpha\downarrow 0} V_2^\alpha(T) = 0$ and $\lim_{\alpha\uparrow\frac{1}{2}} V_2^\alpha(T) = \beta$.

To examine the dependence of b^* on α, we write $b^*(\alpha)$ and ℓ_α instead of b^* and ℓ, respectively. Moreover, we consider the modified equation

$$\frac{\ell_\alpha(b)}{1-2\alpha} = \frac{\beta-1}{1-2\alpha}[q(\alpha)-q(b)+(b-\alpha)q'(b)]+q'(b)=0. \tag{IV.2.2}$$

Then b^* is a solution to $\ell_\alpha(b)=0$ if and only if b^* is a solution to (IV.2.2). Notice that

$$\frac{\partial}{\partial\alpha}\frac{\ell_\alpha(b)}{1-2\alpha} = \frac{\beta-1}{(1-2\alpha)^2}\big\{2[q(\alpha)-q(b)+(b-\alpha)q'(\alpha)]-(1-2b)[q'(b)-q'(\alpha)]\big\}<0 \tag{IV.2.3}$$

for $b \in \left(\alpha, \frac{1}{2}\right]$ by the strict convexity of q. Now let $0<\alpha<\gamma<\frac{1}{2}$. Then (IV.2.3) yields for all $b \in \left(\gamma, \frac{1}{2}\right)$ that

$$\frac{\ell_\alpha(b)}{1-2\alpha} > \frac{\ell_\gamma(b)}{1-2\gamma}. \tag{IV.2.4}$$

In particular, by (IV.2.4) and since $\frac{\ell_\gamma(b)}{1-2\gamma}$ is strictly increasing in b, we have for all $b \geq b^*(\gamma)$ that

$$\frac{\ell_\alpha(b)}{1-2\alpha} > \frac{\ell_\gamma\big(b^*(\gamma)\big)}{1-2\gamma}=0.$$

Therefore, $b^*(\alpha)<b^*(\gamma)$, i.e. b^* is strictly increasing in α.

Since $\frac{\frac{1}{2}-a}{1-\alpha-a}q(\alpha)+\frac{\frac{1}{2}-\alpha}{1-\alpha-a}q(a)$ strictly decreases in a for fixed α and in α for fixed a, it follow that T^* decreases in α. Indeed,

$$\begin{aligned}
T^*(\alpha) &= \frac{\frac{1}{2}-b^*(\alpha)}{1-\alpha-b^*(\alpha)}q(\alpha)+\frac{\frac{1}{2}-\alpha}{1-\alpha-b^*(\alpha)}q\big(b^*(\alpha)\big)\\
&> \frac{\frac{1}{2}-b^*(\gamma)}{1-\alpha-b^*(\gamma)}q(\alpha)+\frac{\frac{1}{2}-\alpha}{1-\alpha-b^*(\gamma)}q\big(b^*(\gamma)\big)\\
&> \frac{\frac{1}{2}-b^*(\gamma)}{1-\gamma-b^*(\gamma)}q(\alpha)+\frac{\frac{1}{2}-\gamma}{1-\gamma-b^*(\gamma)}q\big(b^*(\gamma)\big)=T^*(\gamma).
\end{aligned}$$

Dependence on β

Observe that q and $a^*(T)$ are independent of β. Since the payoff function f is increasing in β, the functions V_2 and V increase in β. Furthermore, $V_2(T)$ is linear in β, thus, it strictly increases in β with $\lim_{\beta\to\infty} V_2(T)=\infty$ for $T\in(0,\infty)$ and for $\beta=1$ it holds that

$$V_2(T)=\begin{cases}\dfrac{\frac{1}{2}-a^*(T)}{1-\alpha-a^*(T)}, & T\in\big[0,q(\alpha)\big),\\ 1, & T\geq q(\alpha).\end{cases}$$

For the dependence of b^* on β we use that

$$\ell(b)=(\beta-1)\big(q(\alpha)-q(b)\big)+(1-\alpha-b+b\beta-\alpha\beta)q'(b)$$

strictly increases in b for fixed β as well as in β for fixed $b\in\left(\alpha,\frac{1}{2}\right]$. Therefore, b^* strictly decreases in β. Furthermore, we have $b^*=\frac{1}{2}$ if $\beta=1$ and $b^*\downarrow\alpha$ for $\beta\to\infty$. To see the second claim, first note that q' strictly increases on $\left[\alpha,\frac{1}{2}\right]$ with $q'\left(\frac{1}{2}\right)=0$. Thus, for $\beta>\frac{1}{2\alpha}$ it holds that

$$\begin{aligned}
\ell(b) &= (\beta-1)\big[q(\alpha)-q(b)+(b-\alpha)q'(b)\big]+(1-2\alpha)q'(b)\\
&> (\beta-1)\left[q(\alpha)-q(b)+(b-\alpha)q'(b)+\frac{q'(\alpha)}{\beta}\right].
\end{aligned} \tag{IV.2.5}$$

Using that $r(b) := q(\alpha) - q(b) + (b-\alpha)q'(b)$ is strictly increasing on $\left[\alpha, \frac{1}{2}\right]$ and continuous with $r(\alpha) = 0$ and $r\left(\frac{1}{2}\right) = q(\alpha)$, we conclude that for $\beta > \max\left\{\frac{-q'(\alpha)}{q(\alpha)}, \frac{1}{2\alpha}\right\}$ there exists $d(\beta) := r^{-1}\left(\frac{-q'(\alpha)}{\beta}\right) \in \left(\alpha, \frac{1}{2}\right)$ such that

$$r\big(d(\beta)\big) + \frac{q'(\alpha)}{\beta} = 0.$$

In particular, for all $b > d(\beta)$ we have $r(b) > -\frac{q'(\alpha)}{\beta} > 0$, which together with (IV.2.5) and the fact that ℓ is strictly increasing implies that $b^* = b^*(\beta) \in (\alpha, d(\beta))$. Since $\frac{q'(\alpha)}{\beta} \to 0$ for $\beta \to \infty$, we conclude that $d(\beta) \downarrow \alpha$ and therefore, $b^*(\beta) \downarrow \alpha$ as $\beta \to \infty$.

For the dependence of T^* on β observe that $b \mapsto \frac{\frac{1}{2}-b}{1-\alpha-b}q(\alpha) + \frac{\frac{1}{2}-\alpha}{1-\alpha-b}q(b)$ is decreasing and b^* decreases in β. Hence, it holds that T^* increases in β with $T^* = 0$ if $\beta = 1$ and $\lim_{\beta\to\infty} T^* = q(\alpha)$.

Dependence on the Strength of the Signal $\frac{\kappa}{\sigma}$

First note that the expressions only depend on $\left|\frac{\kappa}{\sigma}\right|$. The function q strictly decreases in $\left|\frac{\kappa}{\sigma}\right|$ on $(0,1)\setminus\left\{\frac{1}{2}\right\}$ and $q\left(\frac{1}{2}\right) = 0$ for all $\left|\frac{\kappa}{\sigma}\right| > 0$. To emphasize the dependence on $\left|\frac{\kappa}{\sigma}\right|$, we write $q^{\kappa/\sigma}$ instead of q in the following. Using that $a^*(T)$, $T \in [0,\infty)$, is the unique solution of

$$T = \frac{\frac{1}{2}-\alpha}{1-\alpha-a}q^{\kappa/\sigma}(a) + \frac{\frac{1}{2}-a}{1-\alpha-a}q^{\kappa/\sigma}(\alpha) = \left|\frac{\sigma}{\kappa}\right|^2 \left(\frac{\frac{1}{2}-\alpha}{1-\alpha-a}q^1(a) + \frac{\frac{1}{2}-a}{1-\alpha-a}q^1(\alpha)\right)$$

we conclude that $a^*(T)$ solves

$$\frac{\frac{1}{2}-\alpha}{1-\alpha-a}q^1(a) + \frac{\frac{1}{2}-a}{1-\alpha-a}q^1(\alpha) = \left|\frac{\kappa}{\sigma}\right|^2 T.$$

Since $\frac{\frac{1}{2}-\alpha}{1-\alpha-a}q^1(a) + \frac{\frac{1}{2}-a}{1-\alpha-a}q^1(\alpha)$ strictly decreases in $a \in \left(0, \frac{1}{2}\right)$, it holds that $a^*(T)$ strictly decreases in $\left|\frac{\kappa}{\sigma}\right|$ for $T \in (0,\infty)$. In addition, it holds that $a^*(T) \to \frac{1}{2}$ for $\left|\frac{\kappa}{\sigma}\right| \to 0$ and $a^*(T) \to 0$ for $\left|\frac{\kappa}{\sigma}\right| \to \infty$.

Using that $a \mapsto \frac{\frac{1}{2}-a}{1-\alpha-a}$ is strictly decreasing on $\left(0, \frac{1}{2}\right)$ we conclude that $V_2(T)$ is strictly increasing in $\left|\frac{\kappa}{\sigma}\right|$. The point b^* is independent of the ratio $\frac{\kappa}{\sigma}$, because

$$(\beta-1)\left[q^{\kappa/\sigma}(\alpha) - q^{\kappa/\sigma}(b)\right] + (1-\alpha-b+\beta b-\alpha\beta)\left(q^{\kappa/\sigma}\right)'(b) = 0$$

holds if and only if

$$(\beta-1)\left[q^1(\alpha) - q^1(b)\right] + (1-\alpha-b+\beta b-\alpha\beta)\left(q^1\right)'(b) = 0.$$

For T^* we again use $q^{\kappa/\sigma} = \left|\frac{\sigma}{\kappa}\right|^2 q^1$ to obtain

$$T^* = \left|\frac{\sigma}{\kappa}\right|^2 \left(\frac{\frac{1}{2}-\alpha}{1-\alpha-b^*}q^1(b^*) + \frac{\frac{1}{2}-b^*}{1-\alpha-b^*}q^1(\alpha)\right).$$

Therefore, T^* is strictly decreasing in $\left|\frac{\kappa}{\sigma}\right|$ with $\lim_{|\kappa/\sigma|\to 0} T^* = \infty$ and $\lim_{|\kappa/\sigma|\to\infty} T^* = 0$. Similarly, we deduce that $q^{\kappa/\sigma}(\alpha) - T^*$ decreases in $\left|\frac{\kappa}{\sigma}\right|$.

The auxiliary stopping point b_1 is given by

$$\begin{aligned} b_1 &= 1-\alpha - \frac{(1-2\alpha)(1-\alpha-b^*)\big(q(\alpha)-T\big)}{\big(\frac{3}{2}-2\alpha-b^*\big)q(\alpha) - \big(\frac{1}{2}-\alpha\big)q(b^*) - (1-\alpha-b^*)T} \\ &= 1-\alpha - \frac{(1-2\alpha)(1-\alpha-b^*)\big[q^1(\alpha) - q^1\big(a^*(T)\big)\big]}{\big(2-2\alpha-b^*-a^*(T)\big)q^1(\alpha) - \big(1-\alpha-a^*(T)\big)q^1(b^*) - (1-\alpha-b^*)q^1\big(a^*(T)\big)}, \end{aligned}$$

where we use that $a^*(T)$ solves $\frac{\frac{1}{2}-\alpha}{1-\alpha-a}q(a)+\frac{\frac{1}{2}-a}{1-\alpha-a}q(\alpha)=T$. Since b^* is independent of $\left|\frac{\kappa}{\sigma}\right|$, we consider

$$\frac{\partial}{\partial a}\left(-\frac{(1-2\alpha)(1-\alpha-b^*)\big[q^1(\alpha)-q^1(a)\big]}{(2-2\alpha-b^*-a)q^1(\alpha)-(1-\alpha-a)q^1(b^*)-(1-\alpha-b^*)q^1(a)}\right)$$

$$=-\frac{(1-2\alpha)(1-\alpha-b^*)\big(q^1(\alpha)-q^1(b^*)\big)\big[q^1(\alpha)-q^1(a)-(1-\alpha-a)\left(q^1\right)'(a)\big]}{\big[(2-2\alpha-b^*-a)q^1(\alpha)-(1-\alpha-a)q^1(b^*)-(1-\alpha-b^*)q^1(a)\big]^2}<0.$$

Using that $a^*(T)$ decreases in $\left|\frac{\kappa}{\sigma}\right|$, the stopping point b_1 is increasing.

IV.3 Optimal Stopping Rules for general a priori Distributions

In this section we state the main results for all a priori probabilities $y\in(0,1)$ of $\{\theta=1\}$. Hence, the a posteriori probability process $(Y_t)_{t\in[0,\infty)}$ starts in $y\in(0,1)$. The statements are proven in Section IV.4.

Lemma IV.3.1. *The value function of the optimal stopping problem* (IV.1.3) *is given by*

$$V_2(T,y)=\begin{cases}\dfrac{c^*(T,y)-y}{c^*(T,y)-\alpha}, & y\in(\alpha,b^*],\ T\le T^*(y),\\[2ex] \beta\dfrac{y-a^*(T,y)}{1-\alpha-a^*(T,y)}, & y\in(b^*,1-\alpha),\ T\le T^*(y),\\[2ex] \max\left\{\dfrac{c^*(T,y)-y}{c^*(T,y)-\alpha},\beta\dfrac{y-a^*(T,y)}{1-\alpha-a^*(T,y)}\right\}, & y\in(\alpha,1-\alpha),\, T\in\big(T^*(y),q(\alpha)-q(y)\big),\\[2ex] 1+(\beta-1)\dfrac{y-a^*(T,y)}{1-\alpha-a^*(T,y)}, & y\in(0,\alpha]\ \textit{or}\\ & y\in(\alpha,1-\alpha),\ T\ge q(\alpha)-q(y),\\[2ex] \beta, & y\ge 1-\alpha,\end{cases}$$

where

- $b^*\in\left(\alpha,\frac{1}{2}\right]$ *is the unique solution on* $[\alpha,1-\alpha]$ *of*
$$\ell(b):=(\beta-1)\big(q(\alpha)-q(b)\big)+(1-\alpha-b+b\beta-\alpha\beta)q'(b)=0, \tag{IV.3.1}$$
- $a^*(T,y)$ *is the unique solution of* $\frac{1-\alpha-y}{1-\alpha-a}q(a)+\frac{y-a}{1-\alpha-a}q(\alpha)-q(y)=T$ *on* $(0,y]$,
- $c^*(T,y)$ *is the unique solution of* $\frac{c-y}{c-\alpha}q(\alpha)+\frac{y-\alpha}{c-\alpha}q(c)-q(y)=T$ *on* $[y,1)$,
- $$T^*(y)=\begin{cases}\dfrac{(b^*-y)q(\alpha)+(y-\alpha)q(b^*)}{b^*-\alpha}-q(y), & y\in(\alpha,b^*],\\[2ex] \dfrac{(1-\alpha-y)q(b^*)+(y-b^*)q(\alpha)}{1-\alpha-b^*}-q(y), & y\in(b^*,1-\alpha].\end{cases} \tag{IV.3.2}$$

Theorem IV.3.2. *The value function V of the optimal stopping problem* (IV.1.2) *is given by*

$$V(T,y)=\begin{cases}\dfrac{T+q(y)-q(b^*)}{q(\alpha)-q(b^*)}+(\beta-1)\left(\dfrac{y-\alpha}{1-2\alpha}-\dfrac{(b^*-\alpha)\big(q(\alpha)-T-q(y)\big)}{(1-2\alpha)\big(q(\alpha)-q(b^*)\big)}\right), & \\ & y\in(\alpha,1-\alpha),\ T\in\big(T^*(y),q(\alpha)-q(y)\big),\\ V_2(T,y), & \textit{else}.\end{cases}$$

An optimal stopping time τ^* *for* $V(T,y)$ *is given by*

$$\tau^* = \begin{cases} \tau\big(\alpha, c^*(T,y)\big), & y \in (\alpha, b^*],\ T \le T^*(y), \\[1ex] \tau\big(a^*(T,y), 1-\alpha\big), & y \in (0,\alpha] \text{ or} \\ & y \in (\alpha, b^*],\ T \ge q(\alpha) - q(y) \text{ or} \\ & y \in (b^*, 1-\alpha),\ T \in [0, T^*(y)] \cup [q(\alpha)-q(y), \infty), \\[1ex] 0, & y \ge 1-\alpha. \end{cases}$$

For $y \in (\alpha, 1-\alpha)$ *and* $T \in \big(T^*(y), q(\alpha) - q(y)\big)$ *the stopping time*

$$\tau^* = \tau(\alpha, b_1) + \mathbb{1}_{\{Y_{\tau(\alpha,b_1)} = b_1\}} \inf\big\{t \in [0,\infty) \colon Y_{\tau(\alpha,b_1)+t} \notin (b^*, 1-\alpha)\big\}$$

is optimal in (IV.1.2)*, where*

$$b_1 = 1-\alpha - \frac{(1-2\alpha)(1-\alpha-b^*)\big(q(\alpha) - T - q(y)\big)}{(1-2\alpha-b^*+y)q(\alpha) - (y-\alpha)q(b^*) - (1-\alpha-b^*)\big(T + q(y)\big)} \in (b^* \vee y, 1-\alpha).$$

In particular, for $y \in (\alpha, 1-\alpha)$ *and* $T \in \big(T^*(y), q(\alpha)-q(y)\big)$ *the law of* Y_{τ^*} *is purely atomic with mass points* α, b^* *and* $1-\alpha$ *that do not depend on* T *nor* y*.*

Remark IV.3.3. The value function $V_2(T,y)$ is not continuous in $\big(q(\alpha)-q(y), y\big)$ for $y \in (\alpha, 1-\alpha)$. Indeed, observe that $a^*(T,y)$ and $c^*(T,y)$ are continuous in T for fixed y, because they are the inverse functions of a strictly decreasing respectively increasing and continuous function. Furthermore, we have $a^*\big(\overline{T}, y\big) = \alpha$ and $c^*\big(\overline{T}, y\big) = 1-\alpha$, where $\overline{T} = q(\alpha) - q(y)$. Hence,

$$\lim_{T \uparrow \overline{T}} V_2(T,y) = \max\left\{\frac{1-\alpha-y}{1-2\alpha}, \beta\frac{y-\alpha}{1-2\alpha}\right\} < \frac{1-\alpha-y}{1-2\alpha} + \beta\frac{y-\alpha}{1-2\alpha} = V_2\big(\overline{T}, y\big).$$

Remark IV.3.4. We can calculate $a^*(T,y)$, $c^*(T,y)$ and b^* numerically using Newton's method for given $\alpha \in \big(0, \frac{1}{2}\big)$, $\beta \ge 1$, $y \in (0, 1-\alpha)$ and $T \in [0,\infty)$.

In the following we state that for some average time constraints the expected payoff increases if we allow not only for two possible outcomes but for a third possibility, i.e. the law of the process at the stopping time has three mass points.

Corollary IV.3.5. *Let* $y \in (\alpha, 1-\alpha)$ *and* $T \in \big(T^*(y), q(\alpha)-q(y)\big)$*, where* $T^*(y)$ *is defined in* (IV.3.2)*. Then*

$$V_2(T,y) < \sup\{\mathbb{E}^y[f(Y_\tau)] \colon \mathbb{E}^y[\tau] \le T,\ Y_\tau \text{ has at most 3 mass points}\} = V(T,y).$$

Moreover, V solves the PDE (II.4.6).

Theorem IV.3.6. *The value function* V *given in Theorem IV.3.2 is a classical solution to the PDE*

$$\min\left\{V_T(T,y), V_T(T,y) - \frac{1}{2}\frac{\kappa^2}{\sigma^2}y^2(1-y)^2 V_{yy}(T,y) + \frac{\frac{\kappa^2}{\sigma^2}y^2(1-y)^2 V_{Ty}^2(T,y)}{2V_{TT}(T,y)}\right\} = 0 \qquad \text{(IV.3.3)}$$

on $\big((0,\infty)\times(0,1)\big)\backslash B_\beta$ *with* $V(0,y) = f(y)$*,* $y \in (0,1)$*, where*

$$B_\beta = \begin{cases} \big\{\big(q(\alpha)-q(y), y\big) \colon y \in (\alpha, 1-\alpha)\big\} \cup \big\{\big(T^*(y), y\big) \colon y \in (\alpha, 1-\alpha)\big\} & \beta > 1, \\ \qquad \cup\, (0,\infty) \times \{1-\alpha\}, & \\[1ex] \big\{\big(q(\alpha)-q(y), y\big) \colon y \in (\alpha, 1-\alpha)\big\} \cup \big\{\big(T^*(y), y\big) \colon y \in (\alpha, 1-\alpha)\big\}, & \beta = 1. \end{cases}$$

Recall that we use the convention that the fraction V_{Ty}^2/V_{TT} *is set to zero if both the numerator and denominator equal* 0*.*

IV.4 Proof of Lemma IV.3.1, Theorem IV.3.2, IV.3.6 and Corollary IV.3.5

We show Lemma IV.3.1 and then use its statement to prove Theorem IV.3.2. Corollary IV.3.5 follows from the proof of Theorem IV.3.2. Finally, we show that the value function V satisfies the PDE (IV.3.3).

First we reduce the optimal stopping problems (IV.1.2) and (IV.1.3) to measure optimization problems using the results of Chapter III. Here $\mathcal{A}\,(T, y)$ denotes the set of all probability measures with support in $(0,1)$ such that $\int_{\mathbb{R}} x\,\mu(dx) = y$ and $\int_{\mathbb{R}} q_y(x)\mu(dx) \leq T$. Furthermore, let $\mathcal{A}_n(T,y)$ be the set of all discrete measures in $\mathcal{A}\,(T,y)$ with at most n mass points, $n \in \mathbb{N}$. From Theorem III.2.5 we conclude that

$$V(T,y) = \sup_{\mu \in \mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx), \tag{IV.4.1}$$

$$V_2\,(T,y) = \sup_{\mu \in \mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx). \tag{IV.4.2}$$

In the proof of Lemma IV.3.1 we only focus on V as the value function of an optimal stopping problem. On the other hand, for the proof of Theorem IV.3.2 we use both the characterization of the value function as a measure optimization problem and as an optimal stopping problem. Changing the perspective allows us to simplify and shorten the arguments.

Proof of Lemma IV.3.1. In the following we obtain the value of the optimal stopping problem (IV.1.3). We consider the cases $y \in (0,\alpha]$, $y \in (\alpha, 1-\alpha)$ and $y \in [1-\alpha, 1)$ separately. Furthermore, we assume that $T \in (0,\infty)$, because $V(0,y) = f(y)$ for all $y \in (0,1)$.

$\boxed{y \in [1-\alpha, 1):}$ Since f is bounded above by β, stopping immediately is optimal for $V_2(T,y)$, $T \in (0,\infty)$, and hence $V_2(T,y) = \beta$.

$\boxed{y \in (0,\alpha]:}$ Stopping directly yields a payoff of 1. This is optimal for $\beta = 1$. If $\beta > 1$, then stopping at $1-\alpha$ and $a \in (0,y)$ has a higher payoff. Here a has to be chosen in such a way that the stopping time satisfies the expectation constraint. Using the first exit time of (b,c), $b \in (0,y)$, $c \in (1-\alpha, 1)$, is not optimal, because f is constant on $(0,\alpha]$ and $[1-\alpha, 1)$, respectively. Indeed, once the process attains $1-\alpha$ the agent can obtain a payoff of β, but if she does not stop at $1-\alpha$, then with a positive probability the process goes below $1-\alpha$ and does not return to $1-\alpha$ within the given expected time. Observe that in an optimal strategy the whole time horizon is exploited, because decreasing the point a, which entails a higher expected time for exiting $(a, 1-\alpha)$, increases the probability that the process hits $1-\alpha$ before a. Thus, it is sufficient to focus on $\tau(a, 1-\alpha)$ and choose a such that $\mathbb{E}^y[\tau(a,1-\alpha)] = T$. Lemma 2.2 in [5] implies that

$$\mathbb{E}^y[\tau(a,1-\alpha)] = \mathbb{E}^y\left[q_y\left(Y_{\tau(a,1-\alpha)}\right)\right].$$

The law μ^a of $Y_{\tau(a,1-\alpha)}$ is given by

$$\mu^a = \frac{1-\alpha-y}{1-\alpha-a}\delta_a + \frac{y-a}{1-\alpha-a}\delta_{1-\alpha}.$$

Now we determine a such that

$$\mathbb{E}^y[\tau(a,1-\alpha)] = \int_{\mathbb{R}} q_y(x)\mu^a(dx) = \frac{1-\alpha-y}{1-\alpha-a}q_y(a) + \frac{y-a}{1-\alpha-a}q_y(1-\alpha) = T. \tag{IV.4.3}$$

Using (IV.1.4) and $q(x) = q(1-x)$ we rewrite (IV.4.3) as

$$\int_{\mathbb{R}} q_y(x)\mu^a(dx) = \frac{1-\alpha-y}{1-\alpha-a}q(a) + \frac{y-a}{1-\alpha-a}q(\alpha) - q(y) = T. \tag{IV.4.4}$$

The map

$$k_y\colon\ (0,y]\to[0,\infty),\quad a\mapsto\frac{1-\alpha-y}{1-\alpha-a}q(a)+\frac{y-a}{1-\alpha-a}q(\alpha)-q(y),\tag{IV.4.5}$$

is continuous and strictly decreasing, because

$$k_y'(a)=\frac{1-\alpha-y}{(1-\alpha-a)^2}\big(q(a)+(1-\alpha-a)q'(a)-q(\alpha)\big)<0$$

by the strict convexity of q and since $q(\alpha)=q(1-\alpha)$. Moreover, it holds that $\lim_{a\downarrow 0}k_y(a)=\infty$ and $k_y(y)=0$. Hence, for every $T\in(0,\infty)$ there exists a unique $a^*(T,y)=k_y^{-1}(T)\in(0,y)$ such that (IV.4.4) and, thus, (IV.4.3) hold. Therefore,

$$V_2(T,y)=\beta\frac{y-a^*(T,y)}{1-\alpha-a^*(T,y)}.$$

$\boxed{y\in(\alpha,1-\alpha):}$ Note that the expected time until the process $(Y_t)_{t\in[0,\infty)}$ hits either α or $1-\alpha$ for the first time is given by

$$\mathbb{E}^y[\tau(\alpha,1-\alpha)]=\mathbb{E}^y[q_y(Y_\tau)]=\frac{1-\alpha-y}{1-2\alpha}q_y(\alpha)+\frac{y-\alpha}{1-2\alpha}q_y(1-\alpha)=q(\alpha)-q(y).$$

The last equality follows from (IV.1.4). Now we distinguish between the cases $T<q(\alpha)-q(y)$ and $T\geq q(\alpha)-q(y)$.

$\boxed{T<q(\alpha)-q(y):}$ In this case the process cannot reach both α and $1-\alpha$ within the given expected time horizon. Thus, at least one of the stopping points lies inside $(\alpha,1-\alpha)$. Similar to the case where $y\in(0,\alpha]$, we conclude that either stopping at α and $c\in(y,1-\alpha)$ such that $\mathbb{E}^y[\tau(\alpha,c)]=T$ or stopping at the first exit time of $(a,1-\alpha)$, $a\in(\alpha,y)$, with $\mathbb{E}^y[\tau(a,1-\alpha)]=T$ is optimal. Here we cannot directly argue that the stopping rule $\tau(a,1-\alpha)$ has a higher payoff than $\tau(\alpha,c)$, because if T is small and y is close to α, then it may be better to stop at α than to wait until the process hits $1-\alpha$ with a small probability.

We now derive the stopping points a and c. As in the case $y\in(0,\alpha]$ we conclude that there exists a unique $a^*(T,y)=k_y^{-1}(T)\in(0,y)$ such that $\mathbb{E}^y\left[\tau\big(a^*(T,y),1-\alpha\big)\right]=T$. Furthermore, since k_y is strictly decreasing and $T<q(\alpha)-q(y)=k_y(\alpha)$, it follows that $a^*(T,y)\in(\alpha,y)$. For the stopping point c observe that the map

$$g_y\colon\ [y,1)\to[0,\infty),\quad c\mapsto\frac{c-y}{c-\alpha}q(\alpha)+\frac{y-\alpha}{c-\alpha}q(c)-q(y),\tag{IV.4.6}$$

is continuous and strictly increasing with $g_y(y)=0$, $\lim_{c\uparrow 1}g_y(c)=\infty$ and $g_y(1-\alpha)=q(\alpha)-q(y)$. Therefore, there exists a unique $c^*(T,y)=g_y^{-1}(T)\in(y,1-\alpha)$ such that $\mathbb{E}^y\left[\tau\big(\alpha,c^*(T,y)\big)\right]=T$. To sum up, we have

$$V_2(T,y)=\max\left\{\frac{c^*(T,y)-y}{c^*(T,y)-\alpha},\ \beta\frac{y-a^*(T,y)}{1-\alpha-a^*(T,y)}\right\}.$$

Let $y\in(\alpha,b^*]$ and $T\leq T^*(y)=g_y(b^*)$, where b^* is the unique solution of (IV.3.1) on $[\alpha,1-\alpha]$ (for existence and uniqueness see Lemma A.3.2 in the appendix). In the following we write a^* and c^* instead of $a^*(T,y)$ and $c^*(T,y)$, respectively. We show that for $T\leq T^*(y)$ it holds that

$$V_2(T,y)=\frac{c^*-y}{c^*-\alpha}\geq\beta\frac{y-a^*}{1-\alpha-a^*}\tag{IV.4.7}$$

with strict inequality if $\beta>1$. Assume that

$$\beta\frac{y-a}{1-\alpha-a}>\frac{c^*-y}{c^*-\alpha}\tag{IV.4.8}$$

holds for $a \in (\alpha, y)$. Then $a < y - \frac{(1-\alpha-y)(c^*-y)}{\beta(c^*-\alpha)-(c^*-y)} =: \overline{a} = \overline{a}(T, y)$, because $\frac{y-a}{1-\alpha-a}$ is strictly decreasing in a and $\overline{a}$ is chosen in such a way that equality holds in (IV.4.8) when replacing a by $\overline{a}$. Observe that $\overline{a} \in (\alpha, y)$ if and only if $\beta > \frac{(1-2\alpha)(c^*-y)}{(y-\alpha)(c^*-\alpha)}$. If $\overline{a} \leq \alpha$, then (IV.4.8) does not hold for any $a \in (\alpha, y)$ and hence (IV.4.7) follows. If $\overline{a} \in (\alpha, y)$, then we show that $k_y(a) > g_y(c^*) = T$ for all $a \in (\alpha, \overline{a})$. In particular, we conclude that $a^* \geq \overline{a}$ and as a consequence (IV.4.7) holds.

Since $T \leq T^*(y) = g_y(b^*)$ and g_y is strictly increasing, it follows that $c^* = g_y^{-1}(T) \leq b^*$. First assume that $\beta > 1$. Since ℓ is strictly increasing and we have $c^* \leq b^*$, it follows that $\ell(c^*) < \ell(b^*) = 0$ and thus,

$$q(\alpha) < q(c^*) - \frac{1 - \alpha + c^* - c^*\beta - \alpha\beta}{\beta - 1} q'(c^*). \tag{IV.4.9}$$

The definition of $\overline{a}$, (IV.4.9) and the strict convexity of q imply that

$$\begin{aligned}
k_y(\overline{a}) - g_y(c^*) &= -(\beta - 1)\frac{y - \overline{a}}{1 - \alpha - \overline{a}} q(\alpha) + \frac{1 - \alpha - y}{1 - \alpha - \overline{a}} q(\overline{a}) - \left(1 - \beta \frac{y - \overline{a}}{1 - \alpha - \overline{a}}\right) q(c^*) \\
&> \frac{1 - \alpha - y}{1 - \alpha - \overline{a}} \big[q(\overline{a}) - q(c^*)\big] + \frac{y - \overline{a}}{1 - \alpha - \overline{a}} \big(1 - \alpha - c^* + c^*\beta - \alpha\beta\big) q'(c^*) \\
&> \left(\frac{1 - \alpha - y}{1 - \alpha - \overline{a}} (\overline{a} - c^*) + \frac{y - \overline{a}}{1 - \alpha - \overline{a}} (1 - \alpha - c^* + c^*\beta - \alpha\beta)\right) q'(c^*) \\
&= 0.
\end{aligned}$$

If $\beta = 1$, then

$$k_y(\overline{a}) - g_y(c^*) = \frac{y - \alpha}{c^* - \alpha} [q(\overline{a}) - q(c^*)] \geq \frac{y - \alpha}{c^* - \alpha} (\overline{a} - c^*) q'(c^*) \geq 0,$$

where we use that $\overline{a} < y < c^* \leq b^* = \frac{1}{2}$ and hence $q'(c^*) \leq 0$. Since k_y is strictly decreasing, it follows that $k_y(a) > k_y(\overline{a}) \geq g_y(c^*) = T$ for all $a \in (\alpha, \overline{a})$. Since $a^* = k_y^{-1}(T)$, we conclude that $a^* \geq \overline{a}$ and hence (IV.4.7) holds. In particular, we have $V_2(T, y) = \frac{c^*-y}{c^*-\alpha}$ for $T \leq T^*(y)$. Similarly one can show that

$$V_2(T, y) = \beta \frac{y - a^*(T, y)}{1 - \alpha - a^*(T, y)}$$

for $y \in (b^*, 1 - \alpha)$ and $T \leq T^*(y)$.

$\boxed{T \geq q(\alpha) - q(y)}$ In this case the expected time to reach α and $1 - \alpha$ is less or equal to T. For $T = q(\alpha) - q(y)$ the agent maximizes her payoff by stopping at α and $1 - \alpha$. If $\beta > 1$ and if T increases, then she can increase the probability that the process attains $1 - \alpha$ before hitting a lower bound $a \in [a^*(T, y), \alpha]$ by decreasing a. If she does so, the payoff increases. Since she gains more from increasing the probability at $1 - \alpha$ than at α, stopping at $a^*(T, y)$ and $1 - \alpha$ is optimal. Therefore,

$$V_2(T, y) = \frac{1 - \alpha - y}{1 - \alpha - a^*(T, y)} + \beta \frac{y - a^*(T, y)}{1 - \alpha - a^*(T, y)}.$$

□

Proof of Theorem IV.3.2. First observe that $V(T, y) \geq V_2(T, y)$ for all $T \in [0, \infty)$ and $y \in (0, 1)$. In this proof we also distinguish between the cases $y \in (0, \alpha]$, $y \in (\alpha, 1 - \alpha)$ and $y \in [1 - \alpha, 1)$. Without loss of generality, we assume that $T \in (0, \infty)$.

$\boxed{y \in [1 - \alpha, 1):}$ Since f is bounded above by β, the stopping time $\tau^* = 0$ is optimal. Hence $V(T, y) = V_2(T, y)$ for $T \in (0, \infty)$.

$\boxed{y \in (0,\alpha]:}$ If $\beta = 1$, then we conclude as in the case $y \in [1-\alpha, 1)$ that stopping directly is optimal with $V(T,y) = V_2(T,y) = 1$. If $\beta > 1$, then we use the reformulations (IV.4.1) and (IV.4.2) as well as the fact that V is the value function of the optimal stopping problem (IV.1.2). To embed a measure μ with exactly three mass points $a < b < c$ the agent uses the following stopping rule τ (cf. the proof of Theorem III.1.5): She stops at the first exit time of (a,d), where $d = \big(\mu(\{b\})b + \mu(\{c\})c\big)/\big(\mu(\{b\}) + \mu(\{c\})\big)$ is an auxiliary stopping point. If Y attains a before d, then she is done; otherwise, she uses a second exit time, namely, she waits until the process equals b or c.

From (IV.4.1) we know that stopping at 3 points is enough. Let $\mu \in \mathcal{A}_3(T,y)$ with mass points $a < b < c \in (0,1)$, all having positive probability. Observe that it is not optimal if $a,b,c \in (0,1-\alpha)$, because then the payoff is less or equal to 1, which is strictly smaller than $V_2(T,y)$. Moreover, two mass points in the interval $[1-\alpha,1)$ cannot be optimal neither, because thinking in terms of the optimal stopping problem, the agent also gains β by stopping at $1-\alpha$ but with a higher probability. Stopping at a and $1-\alpha$ has a smaller expected value than stopping at a, b and c where $b, c \geq 1-\alpha$ and hence the stopping time $\tau(a, 1-\alpha)$ is admissible and yields a higher payoff than τ. Similar to the proof of Lemma IV.3.1 we conclude that in an optimal measure it holds that $c = 1-\alpha$. If $b \in (y, 1-\alpha)$, then, compared to the optimal strategy for $V_2(T,y)$, the probability to attain a payoff of β is smaller. Indeed, for the payoff β the process first has to hit the auxiliary stopping point d and then attain $1-\alpha$. But if the agent additionally stops at b after hitting d, the probability of hitting $1-\alpha$ decreases. Hence, $b \in (y, 1-\alpha)$ cannot be optimal. Finally, if $b \in (a, y]$, let

$$\lambda = \frac{(d-y)(1-\alpha-b)}{(d-y)(1-\alpha-b) + (y-a)(1-\alpha-d)}$$

and observe that $\lambda \in (0,1)$. Then stopping at the two points $r = \lambda a + (1-\lambda)b \in (a,y)$ and $1-\alpha$ has the same payoff than using the stopping time associated to μ. Indeed, λ is chosen such that $\mu(\{1-\alpha\}) = \mathbb{P}^y\big[Y_{\tau(r,1-\alpha)} = 1-\alpha\big]$. Hence, the payoff of the two stopping rules coincides. Moreover, it holds that $\big(1-\mu(\{1-\alpha\})\big)\lambda = \mu(\{a\})$ and $\big(1-\mu(\{1-\alpha\})\big)(1-\lambda) = \mu(\{b\})$. The strict convexity of q_y implies that

$$\begin{aligned}
\mathbb{E}^y[\tau(r,1-\alpha)] &= \mathbb{E}^y[q_y(Y_{\tau(r,1-\alpha)})] \\
&= \mathbb{P}^y[Y_{\tau(r,1-\alpha)} = 1-\alpha]q_y(\alpha) + \big(1-\mathbb{P}^y[Y_{\tau(r,1-\alpha)} = 1-\alpha]\big)q_y(r) \\
&< \mu(\{1-\alpha\})q_y(\alpha) + \big(1-\mu(\{1-\alpha\})\big)\big(\lambda q_y(a) + (1-\lambda)q_y(b)\big) \\
&= \int_{\mathbb{R}} q_y(x)\mu(dx) \leq T.
\end{aligned}$$

Therefore, the stopping time $\tau(r, 1-\alpha)$ is an admissible stopping time.

To sum up, this shows that using 3 points instead of 2 does not increase the value of the measure optimization problem (IV.4.1) and thus, $V(T,y) = V_2(T,y)$.

$\boxed{y \in (\alpha, 1-\alpha):}$ Again we examine the cases $T < q(\alpha) - q(y)$ and $T \geq q(\alpha) - q(y)$ separately.

$\boxed{T \geq q(\alpha) - q(y):}$ Recall that $\mathbb{E}^y[\tau(\alpha, 1-\alpha)] = q(\alpha) - q(y)$. Therefore, both points α and $1-\alpha$ can be reached within the given time. Similar arguments as in the case $y \in (0,\alpha]$ show that $V(T,y) = V_2(T,y)$.

$\boxed{T < q(\alpha) - q(y):}$ Here the expected time constraint is to small to reach α and $1-\alpha$ within the given time horizon. Now we use the reformulation (IV.4.1) and consider measures $\mu \in \mathcal{A}_3(T,y) \backslash \mathcal{A}_2(T,y)$. One can argue similarly to the case $y \in (0,\alpha]$ and conclude that it is sufficient to focus on measures μ with mass points $\alpha < b < 1-\alpha$ and to use the full time horizon, i.e. $\int_{\mathbb{R}} q_y(x)\mu(dx) = T$. Note that the line of arguments interprets V as the value of an optimal stopping problem. For every signed measure μ which is atomic and has three mass points, the constraints $\int_{\mathbb{R}} 1\,\mu(dx) = 1$,

$\int_{\mathbb{R}} x\,\mu(dx) = y$ and $\int_{\mathbb{R}} q_y(x)\mu(dx) = T$ uniquely define the weights μ.To obtain a probability measure, one has to restrict the mass points to certain intervals. Let

$$\mu^b = \frac{(1-\alpha-b)\big(T+q(y)\big)+(b-y)q(\alpha)-(1-\alpha-y)q(b)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)}\delta_\alpha + \frac{q(\alpha)-\big(T+q(y)\big)}{q(\alpha)-q(b)}\delta_b$$
$$+\frac{(b-\alpha)\big(T+q(y)\big)-(b-y)q(\alpha)-(y-\alpha)q(b)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)}\delta_{1-\alpha},$$

where $b\in\big(k_y^{-1}(T), g_y^{-1}(T)\big)$ and the functions k_y and g_y are given by (IV.4.5) and (IV.4.6), respectively. By Lemma A.3.1 in the appendix the measures μ^b, $b\in\big(k_y^{-1}(T), g_y^{-1}(T)\big)$, are exactly all measures in $\mathcal{A}_3(T,y)\backslash\mathcal{A}_2(T,y)$ such that the mass is concentrated in α, b and $1-\alpha$ and $\int_{\mathbb{R}} q_y(x)\mu^b(dx) = T$. We now maximize $\int_{\mathbb{R}} f(x)\mu^b(dx)$ over $b\in\big(k_y^{-1}(T), g_y^{-1}(T)\big)\subseteq(\alpha,1-\alpha)$. Let

$$n_y(b) = \int_{\mathbb{R}} f(x)\mu^b(dx) = \frac{T+q(y)-q(b)}{q(\alpha)-q(b)} + (\beta-1)\left(\frac{y-\alpha}{1-2\alpha} - \frac{(b-\alpha)\big(q(\alpha)-T-q(y)\big)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)}\right).$$

Note that

$$n_y'(b) = -\frac{q(\alpha)-T-q(y)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)^2}\left[(\beta-1)\big(q(\alpha)-q(b)\big)+(1-\alpha-b+b\beta-\alpha\beta)q'(b)\right].$$

Since $T<q(\alpha)-q(y)$ we conclude that for all $b\in\big(k_y^{-1}(T), g_y^{-1}(T)\big)$

$$-\frac{q(\alpha)-T-q(y)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)^2}<0.$$

Define the function $\ell\colon [\alpha,1-\alpha]\to\mathbb{R}$ by

$$\ell(b) = (\beta-1)\big(q(\alpha)-q(b)\big)+(1-\alpha-b+b\beta-\alpha\beta)q'(b).$$

Lemma A.3.2 in the appendix implies that ℓ is strictly increasing and that there exists a unique $b^*\in\big(\alpha,\frac{1}{2}\big]$ such that $\ell(b^*)=0$. If $b^*\in\big(k_y^{-1}(T), g_y^{-1}(T)\big)$, then b^* maximizes n_y over $\big(k_y^{-1}(T), g_y^{-1}(T)\big)$. If $b^*<k_y^{-1}(T)$ or $b^*>g_y^{-1}(T)$, then n_y is strictly decreasing respectively increasing on $\big(k_y^{-1}(T), g_y^{-1}(T)\big)$. To examine which conditions guarantee that $b^*\in\big(k_y^{-1}(T), g_y^{-1}(T)\big)$ define

$$T^*(y) = \begin{cases} g_y(b^*) = \dfrac{(b^*-y)q(\alpha)+(y-\alpha)q(b^*)}{b^*-\alpha}-q(y), & y\in(\alpha,b^*],\\[2ex] k_y(b^*) = \dfrac{(1-\alpha-y)q(b^*)+(y-b^*)q(\alpha)}{1-\alpha-b^*}-q(y), & y\in(b^*,1-\alpha].\end{cases}$$

Observe that $T^*(b^*)=0$. Since $q(b^*)<q(\alpha)$ and q is strictly convex, it follows that $T^*(y)\in\big(0,q(\alpha)-q(y)\big)$ for $y\neq b^*$. Let $y\in(\alpha,b^*]$. Using $k_y^{-1}(T)<y$, it holds that $b^*>k_y^{-1}(T)$. Moreover, since g_y is strictly increasing it follows that $b^*<g_y^{-1}(T)$ if and only if $T^*(y)=g_y(b^*)<T$.

$\boxed{T\le T^*(y)\text{:}}$ Then it holds that $b^*>g_y^{-1}(T)$. Thus, n_y is strictly increasing and

$$\begin{aligned}&\sup\left\{n_y(b)\colon b\in\big(k_y^{-1}(T), g_y^{-1}(T)\big)\right\}\\ &= n_y\big(g_y^{-1}(T)\big)\\ &= \frac{T+q(y)-q\big(g_y^{-1}(T)\big)}{q(\alpha)-q\big(g_y^{-1}(T)\big)} + (\beta-1)\left(\frac{y-\alpha}{1-2\alpha} - \frac{\big(g_y^{-1}(T)-\alpha\big)\big(q(\alpha)-T-q(y)\big)}{(1-2\alpha)(q(\alpha)-q\big(g_y^{-1}(T)\big)}\right). \qquad \text{(IV.4.10)}\end{aligned}$$

Recall that $g_y^{-1}(T) = c^*(T, y)$ is the unique solution to $g_y(c) = T$ on $[y, 1)$. Hence, we conclude that

$$q(\alpha) - T - q(y) = \frac{y - \alpha}{c^*(T, y) - \alpha}\big(q(\alpha) - q\big(c^*(T, y)\big)\big), \tag{IV.4.11}$$

$$\frac{T + q(y) - q\big(c^*(T, y)\big)}{q(\alpha) - q\big(c^*(T, y)\big)} = \frac{c^*(T, y) - y}{c^*(T, y) - \alpha}. \tag{IV.4.12}$$

By (IV.4.11) the second summand in (IV.4.10) equals 0. In addition, using (IV.4.12) we conclude that $n_y\big(g_y^{-1}(T)\big)$ simplifies to

$$n_y\big(g_y^{-1}(T)\big) = \frac{T + q(y) - q\big(g_y^{-1}(T)\big)}{q(\alpha) - q\big(g_y^{-1}(T)\big)} = \frac{T + q(y) - q\big(c^*(T, y)\big)}{q(\alpha) - q\big(c^*(T, y)\big)} = \frac{c^*(T, y) - y}{c^*(T, y) - \alpha}.$$

Therefore, if $y \in (\alpha, b^*]$ and $T \leq T^*(y)$, then

$$\sup\left\{\int_{\mathbb{R}} f(x)\mu^b(dx) \colon b \in \big(k_y^{-1}(T), g_y^{-1}(T)\big)\right\} = \frac{c^*(T, y) - y}{c^*(T, y) - \alpha} = V_2(T, y).$$

Thus, $V(T, y) = V_2(T, y)$ and we do not gain more from stopping at 3 points than at 2 points.

$\boxed{T \in \big(T^*(y), q(\alpha) - q(y)\big)\colon}$ In this case it holds that $b^* \in \big(k_y^{-1}(T), g_y^{-1}(T)\big)$ and hence,

$$\begin{aligned}
\sup\left\{\int_{\mathbb{R}} f(x)\mu^b(dx) \colon b \in \big(k_y^{-1}(T), g_y^{-1}(T)\big)\right\} &= n_y(b^*) \\
&= \frac{T + q(y) - q(b^*)}{q(\alpha) - q(b^*)} + (\beta - 1)\left(\frac{y - \alpha}{1 - 2\alpha} - \frac{(b^* - \alpha)\big(q(\alpha) - T - q(y)\big)}{(1 - 2\alpha)\big(q(\alpha) - q(b^*)\big)}\right).
\end{aligned}$$

It remains to show that $V(T, y) = n_y(b^*) > V_2(T, y)$ for $T \in \big(T^*(y), q(\alpha) - q(y)\big)$. In the following we write $a^* = k_y^{-1}(T)$ and $c^* = g_y^{-1}(T)$ instead of $a^*(T, y)$ and $c^*(T, y)$, respectively. (IV.4.11) and (IV.4.12) imply that

$$\begin{aligned}
&n_y(b^*) - \frac{c^* - y}{c^* - \alpha} \\
&= \frac{T + q(y) - q(b^*)}{q(\alpha) - q(b^*)} + (\beta - 1)\left(\frac{y - \alpha}{1 - 2\alpha} - \frac{(b^* - \alpha)\big(q(\alpha) - T - q(y)\big)}{(1 - 2\alpha)\big(q(\alpha) - q(b^*)\big)}\right) - \frac{c^* - y}{c^* - \alpha} \\
&= \frac{y - \alpha}{c^* - \alpha}\left(\frac{q(c^*) - q(b^*)}{q(\alpha) - q(b^*)} + (\beta - 1)\frac{(c^* - b^*)[q(\alpha) - q(b^*)] + (b^* - \alpha)[q(c^*) - q(b^*)]}{(1 - 2\alpha)\big(q(\alpha) - q(b^*)\big)}\right) \\
&= \frac{(y - \alpha)\left\{(c^* - b^*)(\beta - 1)\big(q(\alpha) - q(b^*)\big) + \big(1 - 2\alpha + (\beta - 1)(b^* - \alpha)\big)\big[q(c^*) - q(b^*)\big]\right\}}{(1 - 2\alpha)(c^* - \alpha)\big(q(\alpha) - q(b^*)\big)} \\
&= \frac{(y - \alpha)(1 - \alpha - b^* + b^*\beta - \alpha\beta)\,[q(c^*) - q(b^*) - (c^* - b^*)q'(b^*)]}{(1 - 2\alpha)(c^* - \alpha)\big(q(\alpha) - q(b^*)\big)} \\
&> 0,
\end{aligned}$$

where the last equality follows from $\ell(b^*) = 0$, i.e.

$$(\beta - 1)\big(q(\alpha) - q(b^*)\big) = -(1 - \alpha - b^* + b^*\beta - \alpha\beta)q'(b^*),$$

and the strict convexity of q implies the inequality.

Similar arguments lead to

$$n_y(b^*) - \beta\frac{y-a^*}{1-\alpha-a^*}$$
$$= \frac{(1-\alpha-a^*)(1-\alpha-b^*+b^*\beta-\alpha\beta)\big[q(a^*)-q(b^*)+(b^*-a^*)q'(b^*)\big]}{(1-2\alpha)(1-\alpha-a^*)\big(q(\alpha)-q(b^*)\big)} > 0.$$

Hence,

$$V(T,y) = n_y(b^*) > \max\left\{\frac{c^*-y}{c^*-\alpha}, \beta\frac{y-a^*}{1-\alpha-a^*}\right\} = V_2(T,y)$$

for $y \in (\alpha, b^*]$ and $T \in \big(T^*(y), q(\alpha)-q(y)\big)$. For $y \in (b^*, 1-\alpha)$ we argue similarly to obtain

$$V(T,y) = \begin{cases} V_2(T,y), & T \le T^*(y), \\ \frac{T+q(y)-q(b^*)}{q(\alpha)-q(b^*)} + (\beta-1)\left(\frac{y-\alpha}{1-2\alpha} - \frac{(b^*-\alpha)\big(q(\alpha)-T-q(y)\big)}{(1-2\alpha)\big(q(\alpha)-q(b^*)\big)}\right), & T \in \big(T^*(y), q(\alpha)-q(y)\big). \end{cases}$$

In particular, it holds that $V_2(T,y) < V(T,y)$ for $y \in (b^*, 1-\alpha)$ and $T \in \big(T^*(y), q(\alpha)-q(y)\big)$.

The optimal measure for the reformulated problem (IV.4.1) for $y \in (\alpha, 1-\alpha)$ and $T \in \big(T^*(y), q(\alpha)-q(y)\big)$ is given by

$$\mu^{b^*} = \frac{(1-\alpha-b^*)\big(T+q(y)\big)+(b^*-y)q(\alpha)-(1-\alpha-y)q(b^*)}{(1-2\alpha)\big(q(\alpha)-q(b^*)\big)}\delta_\alpha + \frac{q(\alpha)-T-q(y)}{q(\alpha)-q(b^*)}\delta_{b^*}$$
$$+ \frac{(b^*-\alpha)\big(T+q(y)\big)-(b^*-y)q(\alpha)-(y-\alpha)q(b^*)}{(1-2\alpha)\big(q(\alpha)-q(b^*)\big)}\delta_{1-\alpha}.$$

According to Theorem III.1.5 and its proof, a stopping time τ^*, which embeds μ^{b^*} in $(Y_t)_{t\in[0,\infty)}$ for $y \in (\alpha, 1-\alpha)$ and $T \in \big(T^*(y), q(\alpha)-q(y)\big)$ is given by

$$\tau^* = \tau(\alpha, b_1) + \mathbb{1}_{\{Y_{\tau(\alpha,b_1)}=b_1\}} \inf\big\{t \in [0,\infty)\colon Y_{\tau(\alpha,b_1)+t} \notin (b^*, 1-\alpha)\big\}$$

for $b_1 \in (b^* \vee y, 1-\alpha)$ such that $\mu^{b^*}(\{\alpha\}) = \frac{b_1-y}{b_1-\alpha}$. Hence, we conclude that

$$b_1 = 1-\alpha - \frac{(1-2\alpha)(1-\alpha-b^*)\big(q(\alpha)-T-q(y)\big)}{(1-2\alpha-b^*+y)q(\alpha)-(y-\alpha)q(b^*)-(1-\alpha-b^*)\big(T+q(y)\big)}.$$

Similar arguments apply for $y \in (b^*, 1-\alpha)$ and $T \in \big(T^*(y), q(\alpha)-q(y)\big)$. □

Remark IV.4.1. The optimal measure μ^{b^*} for $V(T,y)$ with $y \in (\alpha, 1-\alpha)$ and $T \in \big(T^*(y), q(\alpha)-q(y)\big)$ concentrates its mass in the points $\alpha, 1-\alpha$ and b^*. In particular, only the probability weights depend on T and y but not the mass points. Hence, the same three points suffice in (IV.4.1).

Proof of Corollary IV.3.5. Follows directly from the proof of Theorem IV.3.2. □

Finally, we prove that V solves the PDE (II.4.6).

Proof of Theorem IV.3.6. We first show how to differentiate $a^*(T,y)$ and $c^*(T,y)$ with respect to T and y. Let $\varphi \in \mathcal{C}^{1,1}\big(A \times Y, (0,\infty)\big)$ for $A, Y \subseteq \mathbb{R}$ open and assume that for fixed $y \in Y$ the mapping $\varphi_y\colon A \to [0,\infty), a \mapsto \varphi(a,y)$, is bijective. Then for $T \in (0,\infty)$ it holds that

$$\frac{\partial}{\partial T}\varphi_y^{-1}(T) = \frac{1}{\varphi_y'\big(\varphi_y^{-1}(T)\big)}.$$

For the partial derivative with respect to y observe that

$$\begin{aligned}0 &= \frac{\mathbf{d}}{\mathbf{d}y}\varphi_y\left(\varphi_y^{-1}(T)\right) = \frac{\mathbf{d}}{\mathbf{d}y}\varphi\left(\varphi_y^{-1}(T), y\right)\\ &= \left(\frac{\partial}{\partial a}\varphi\right)\left(\varphi_y^{-1}(T), y\right)\left(\frac{\partial}{\partial y}\varphi_y^{-1}(T)\right) + \left(\frac{\partial}{\partial y}\varphi\right)\left(\varphi_y^{-1}(T), y\right).\end{aligned}$$

Hence, it follows that

$$\frac{\partial}{\partial y}\varphi_y^{-1}(T) = -\frac{\left(\frac{\partial}{\partial y}\varphi\right)\left(\varphi_y^{-1}(T), y\right)}{\varphi_y'\left(\varphi_y^{-1}(T)\right)}.$$

As a consequence,

$$\begin{aligned}\frac{\partial}{\partial T}a^*(T,y) &= -\frac{\big(1-\alpha-a^*(T,y)\big)^2}{(1-\alpha-y)\left[q(\alpha)-q\big(a^*(T,y)\big)-\big(1-\alpha-a^*(T,y)\big)q'\big(a^*(T,y)\big)\right]},\\ \frac{\partial}{\partial y}a^*(T,y) &= \frac{\big(1-\alpha-a^*(T,y)\big)\left[q(\alpha)-q\big(a^*(T,y)\big)-\big(1-\alpha-a^*(T,y)\big)q'(y)\right]}{(1-\alpha-y)\left[q(\alpha)-q\big(a^*(T,y)\big)-\big(1-\alpha-a^*(T,y)\big)q'\big(a^*(T,y)\big)\right]},\\ \frac{\partial}{\partial T}c^*(T,y) &= \frac{\big(c^*(T,y)-\alpha\big)^2}{(y-\alpha)\left[q(\alpha)-q\big(c^*(T,y)\big)+\big(c^*(T,y)-\alpha\big)q'\big(c^*(T,y)\big)\right]},\\ \frac{\partial}{\partial y}c^*(T,y) &= \frac{\big(c^*(T,y)-\alpha\big)\left[q(\alpha)-q\big(c^*(T,y)\big)+\big(c^*(T,y)-\alpha\big)q'(y)\right]}{(y-\alpha)\left[q(\alpha)-q\big(c^*(T,y)\big)+\big(c^*(T,y)-\alpha\big)q'\big(c^*(T,y)\big)\right]}.\end{aligned}$$

Computing the first partial derivatives of V yields that $V \in \mathcal{C}^{1,1}\big((0,\infty)\times(0,1-\alpha)\big)$ if $\beta > 1$ and $V \in \mathcal{C}^{1,1}\big((0,\infty)\times(0,1)\big)$ if $\beta = 1$ but V is not twice continuously differentiable with respect to T nor y in $\big(q(\alpha)-q(y), y\big)$ and $\big(T^*(y), y\big)$, $y \in (\alpha, 1-\alpha)$. Furthermore, it holds that V is a classical solution to the PDE (IV.3.3) on $\big((0,\infty)\times(0,1-\alpha)\big)\backslash B_\beta$, where

$$B_\beta = \begin{cases}\left\{\big(q(\alpha)-q(y), y\big)\colon y\in(\alpha,1-\alpha)\right\}\cup\left\{\big(T^*(y),y\big)\colon y\in(\alpha,1-\alpha)\right\} & \beta>1,\\ \qquad\cup\,(0,\infty)\times\{1-\alpha\}, & \\ \left\{\big(q(\alpha)-q(y), y\big)\colon y\in(\alpha,1-\alpha)\right\}\cup\left\{\big(T^*(y),y\big)\colon y\in(\alpha,1-\alpha)\right\}, & \beta=1,\end{cases}$$

with initial condition $V(0,y) = f(y)$, $y \in (0,1)$. □

Remark IV.4.2. In the proof of Lemma IV.3.1 we show that

$$\max\left\{\frac{c^*(T,y)-y}{c^*(T,y)-\alpha}, \beta\frac{y-a^*(T,y)}{1-\alpha-a^*(T,y)}\right\} = \begin{cases}\dfrac{c^*(T,y)-y}{c^*(T,y)-\alpha}, & y\in(\alpha,b^*],\ T\le T^*(y),\\ \beta\dfrac{y-a^*(T,y)}{1-\alpha-a^*(T,y)}, & y\in[b^*,1-\alpha),\ T\le T^*(y),\end{cases}$$

where b^* is the unique solution of (IV.3.1) and $T^*(y)$ is given by (IV.3.2).

For $T \in \big(T^*(y), q(\alpha)-q(y)\big)$ it is general not so easy to decide which value the maximum attains. Using the derivatives of $a^*(T,y)$ and $c^*(T,y)$ with respect to y (see the proof of Theorem IV.3.6) results in

$$\begin{aligned}&\frac{\partial}{\partial y}\left(\beta\frac{y-a^*(T,y)}{1-\alpha-a^*(T,y)} - \frac{c^*(T,y)-y}{c^*(T,y)-\alpha}\right)\\ &\qquad= \frac{\beta[q'(y)-q'(a^*)]}{q(\alpha)-q(a^*)-(1-\alpha-a^*)q'(a^*)} + \frac{q'(c^*)-q'(y)}{q(\alpha)-q(c^*)+(c^*-\alpha)q'(c^*)} > 0,\end{aligned}$$

where $a^* = a^*(T,y)$ and $c^* = c^*(T,y)$. Here we use that q is strictly convex and $a^*(T,y) < y < c^*(T,y)$. Moreover, for $y = \frac{1}{2}$ it holds that $c^*\left(T, \frac{1}{2}\right) = 1 - a^*\left(T, \frac{1}{2}\right)$ and thus

$$\beta \frac{\frac{1}{2} - a^*\left(T, \frac{1}{2}\right)}{1 - \alpha - a^*\left(T, \frac{1}{2}\right)} - \frac{c^*\left(T, \frac{1}{2}\right) - \frac{1}{2}}{c^*\left(T, \frac{1}{2}\right) - \alpha} = (\beta - 1) \frac{\frac{1}{2} - a^*\left(T, \frac{1}{2}\right)}{1 - \alpha - a^*\left(T, \frac{1}{2}\right)} \geq 0.$$

Therefore, for $y \geq \frac{1}{2}$ we have for all $T \in [0, \infty)$

$$\max \left\{ \frac{c^*(T,y) - y}{c^*(T,y) - \alpha}, \beta \frac{y - a^*(T,y)}{1 - \alpha - a^*(T,y)} \right\} = \beta \frac{y - a^*(T,y)}{1 - \alpha - a^*(T,y)}.$$

V. 2 Points Suffice

When optimally stopping a one-dimensional regular continuous strong Markov process Y with stopping times τ satisfying the expectation constraint $\mathbb{E}^y[\tau] \leq T$, $T \in [0, \infty)$, it is sufficient to consider stopping times τ such that the law of Y_τ is a weighted sum of at most three Dirac measures. The stopping times embedding these laws can be interpreted as consecutive exit times of intervals. In general we cannot restrict the stopping times to first exit times of intervals, cf. Chapter III.

In this chapter we derive a sufficient condition for the payoff function f which allows us to reduce the set of stopping times to first exit times satisfying the expectation constraint. Consequently, the law of the process Y at these stopping times is a weighted sum of at most two Dirac measures. Hence, 2 points suffice.

For this purpose we introduce the concept of Tchebycheff systems in Section V.1. Let $u_0, ..., u_n, f$ be continuous functions on $[a, b]$, $-\infty < a < b < \infty$. The functions $u_0, ..., u_n$ are supposed to be constraint functions for a measure optimization problem with payoff function f. If both $(u_0, ..., u_n)$ and $(u_0, ..., u_n, f)$ form a Tchebycheff system over $[a, b]$, then we can characterize the unique function ς^* (in general not a distribution function) maximizing $\int_a^b f(x)d\varsigma(x)$ among all non-decreasing right-continuous functions ς of bounded total variation satisfying the constraints $\int_a^b u_i(x)d\varsigma(x) = c_i$, $0 \leq i \leq n$ for suitable $c_0, ..., c_n \in \mathbb{R}$ (Section V.1.6).

Since Tchebycheff systems are not so well-established in the literature on optimal stopping problems, in Section V.1 we collect results from Chapter I and II of [38] that allow to identify ς^* and provide detailed proofs based on [38]. In particular, after giving two equivalent formulations of Tchebycheff systems in Section V.1.1, we show how to construct a function $v = \sum_{i=0}^n a_i u_i$, $a_0, ..., a_n \in \mathbb{R}$, with a prescribed set of zeros for a Tchebycheff system $(u_0, ..., u_n)$ in Section V.1.2. In Section V.1.3 we characterize the set of all u-moments $\big(\int_a^b u_i(x)d\varsigma(x)\big)_{0\leq i\leq n}$, where ς is non-decreasing, right-continuous and of bounded total variation, and identify the boundary and interior points in terms of the index in Section V.1.4 and V.1.5. Finally, we characterize ς^* in Section V.1.6.

In Section V.2.1 we consider a one-dimensional regular continuous strong Markov process Y with state space $[a, b]$, $-\infty < a < b < \infty$, and show that the functions $u_0(x) = 1$, $u_1(x) = x$ and $u_2(x) = q_y(x)$, $x \in [a, b]$, $y \in (a, b)$, which describe the constraints in the measure optimization problem of Chapter III, form a Tchebycheff system. Furthermore, we prove necessary and sufficient conditions for the continuous payoff function f such that (u_0, u_1, u_2, f) also constitutes a Tchebycheff system. This allows us to consider only first exit times of intervals in the optimal stopping problem with expectation constraint. We also describe how to deal with Markov processes Y having non-compact support. In Section V.2.2 we illustrate our results for a Brownian motion on $[a, b]$ which is absorbed at a and b.

V.1 Tchebycheff Systems

This section introduces the concept of Tchebycheff systems over a compact interval and is based on Chapter I and II of [38]. Let $u_0, u_1, ..., u_n, f$, $n \in \mathbb{N}$, be continuous functions on the interval $[a, b]$, $-\infty < a < b < \infty$. If $(u_0, u_1, ..., u_n)$ and $(u_0, ..., u_n, f)$ constitute a Tchebycheff system, then for suitable $c = (c_0, ..., c_n) \in \mathbb{R}^{n+1}$ we can characterize the maximizer ς^* of $\int_a^b f(x)\varsigma(dx)$ over all

Stieltjes measure functions ς satisfying the constraint $\int_a^b u_i(x)\varsigma(dx) = c_i$, $0 \leq i \leq n$.

If $n = 2$, then the measure ν^* associated to the maximizer ς^* turns out to be a conical sum of two Dirac measure and ν^* does not depend on f.

Throughout this chapter let $-\infty < a < b < \infty$.

Definition V.1.1. Let $u_0, ..., u_n \colon [a,b] \to \mathbb{R}$ be continuous functions, $n \in \mathbb{N}$. The system $(u_i)_{i=0}^n = (u_0, ..., u_n)$ constitutes a *Tchebycheff system* over $[a,b]$ if the determinants

$$D\begin{pmatrix} 0 & 1 & ... & n \\ x_0 & x_1 & ... & x_n \end{pmatrix} := \det \begin{pmatrix} u_0(x_0) & u_0(x_1) & ... & u_0(x_n) \\ u_1(x_0) & u_1(x_1) & ... & u_1(x_n) \\ \vdots & \vdots & & \vdots \\ u_n(x_0) & u_n(x_1) & ... & u_n(x_n) \end{pmatrix} \tag{V.1.1}$$

are either strictly positive whenever $a \leq x_0 < x_1 < ... < x_n \leq b$ or strictly negative for all $a \leq x_0 < x_1 < ... < x_n \leq b$.

Remark V.1.2. Note that Karlin and Studden in [38] only consider the case with strictly positive determinant in (V.1.1), because whenever the determinant is strictly negative, we obtain a system with strictly positive determinant by multiplying one function u_i, $0 \leq i \leq n$, with -1. Throughout we examine the case of strictly positive as well as strictly negative determinant.

Example V.1.3. Let $u_i \colon [a,b] \to \mathbb{R}$, $u_i(x) = x^i$ for $0 \leq i \leq n$ and $n \in \mathbb{N}$. Then $(u_i)_{i=0}^n$ forms a Tchebycheff system over $[a,b]$, because the determinant in (V.1.1) is given by the Vandermonde determinant. More precisely,

$$D\begin{pmatrix} 0 & 1 & ... & n \\ x_0 & x_1 & ... & x_n \end{pmatrix} = \prod_{0 \leq i < j \leq n} (x_j - x_i) > 0$$

for all increasing sequences $a \leq x_0 < x_1 < ... < x_n \leq b$.

V.1.1 Equivalent Formulations of Tchebycheff Systems

Here we prove equivalent formulations for Tchebycheff systems in terms of the zero set of linear combinations of $u_0, ..., u_n$. First we define *u-polynomials*.

Definition V.1.4. Let $u_0, u_1, ..., u_n \colon [a,b] \to \mathbb{R}$, $n \in \mathbb{N}$. A function $v \colon [a,b] \to \mathbb{R}$ defined by $v(x) = \sum_{i=0}^n a_i u_i(x)$ with $a_i \in \mathbb{R}$, $0 \leq i \leq n$, is called *u-polynomial*. We call a u-polynomial *non-trivial* if $\sum_{i=0}^n a_i^2 > 0$.

Note that if $(u_i)_{i=0}^n$ is a Tchebycheff system, then the functions $u_0, u_1, ..., u_n$ are linearly independent and hence any u-polynomial $v = \sum_{i=0}^n a_i u_i$ is uniquely determined by the coefficients $a_0, a_1, ..., a_n$. Moreover, v is fully defined by its value at $n+1$ distinct points $a \leq x_0 < x_1 < ... < x_n \leq b$, because $a = (a_0, a_1, ..., a_n)$ is the unique solution of

$$\begin{pmatrix} v(x_0) \\ v(x_1) \\ \vdots \\ v(x_n) \end{pmatrix} = \begin{pmatrix} u_0(x_0) & u_1(x_0) & ... & u_n(x_0) \\ u_0(x_1) & u_1(x_1) & ... & u_n(x_1) \\ \vdots & \vdots & & \vdots \\ u_0(x_n) & u_1(x_n) & ... & u_n(x_n) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}. \tag{V.1.2}$$

In particular, any non-trivial u-polynomial has at most n distinct zeros. Notice that a non-trivial u-polynomial with n prescribed zeros $x_0 < x_1 < ... < x_{n-1}$ in $[a,b]$ is given by

$$v(x) = D\begin{pmatrix} 0 & 1 & ... & n-1 & n \\ x_0 & x_1 & ... & x_{n-1} & x \end{pmatrix}.$$

In the following we show that the reverse statement also holds true: Let $(u_i)_{i=0}^n$ by a system of continuous functions on $[a,b]$ such that every non-trivial u-polynomial has at most n distinct zeros, then $(u_i)_{i=0}^n$ is a Tchebycheff system. First notice that for a system $(u_i)_{i=0}^n$ of continuous functions and for any $a \leq x_0 < x_1 < \ldots < x_n \leq b$ the determinants $D\begin{pmatrix} 0 & 1 & \ldots & n \\ x_0 & x_1 & \ldots & x_n \end{pmatrix}$ cannot achieve both positive and negative values without vanishing for some $(x_i)_{i=0}^n \in \mathbb{R}^{n+1}$. Assume that there exist $a \leq x_0 < x_1 < \ldots < x_n \leq b$ such that the determinant in (V.1.1) vanishes. Thus, there exist two different solutions $a = (a_i)_{i=0}^n$, $b = (b_i)_{i=0}^n \in \mathbb{R}^{n+1}$ of (V.1.2). Now define

$$v(x) = \sum_{i=0}^{n} (a_i - b_i) u_i(x).$$

Then v is a non-trivial u-polynomial which vanishes at the $n+1$ different points $x_0, x_1, \ldots, x_n$. This contradicts the assumption that any non-trivial u-polynomial has at most n distinct zeros. As a consequence, the determinant in (V.1.1) is either strictly positive or strictly negative and thus, $(u_i)_{i=0}^n$ forms a Tchebycheff system.

Definition V.1.5. Let $g\colon [a,b] \to \mathbb{R}$ be a continuous function. We denote by $Z(g)$ the number of distinct zeros of g.

Summarizing the previous discussion we obtain the following lemma.

Lemma V.1.6 (Theorem I.4.1. in [38]). *Let $u_0, u_1, \ldots, u_n\colon [a,b] \to \mathbb{R}$ be continuous functions, $n \in \mathbb{N}$. $(u_i)_{i=0}^n$ is a Tchebycheff system over $[a,b]$ if and only if $Z(v) \leq n$ for every non-trivial u-polynomial v.*

Taking into account whether a u-polynomial changes it sign at a zero, we can also characterize a Tchebycheff system.

Definition V.1.7. Let $g\colon [a,b] \to \mathbb{R}$ be a continuous function. A zero $x_0 \in [a,b]$ of g is called isolated if there exists $\varepsilon > 0$ such that $g(x) \neq 0$ for all $x \in [x_0 - \varepsilon, x_0 + \varepsilon] \cap [a,b], x \neq x_0$. An isolated zero $x_0 \in (a,b)$ is called *nonnodal zero* if the function g does not change its sign at x_0. All other zeros including zeros at the endpoints a and b are called *nodal zeros*. Denote by $\widetilde{Z}(g)$ the number of zeros of g, where nodal zeros are counted once and nonnodal zeros twice.

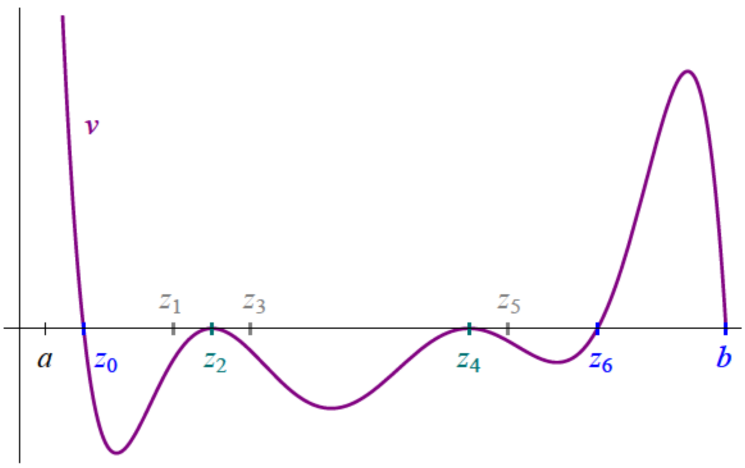


Figure V.1: The zero set of the function v is given by $\mathcal{N} = \{z_0, z_2, z_4, z_6, z_7 = b\}$, where z_2 and z_4 are nonnodal zeros whereas z_0, z_6 and b are nodal zeros. In the proof of Lemma V.1.8 we add the points $z_1 = z_2 - \varepsilon, z_3 = z_2 + \varepsilon$ and $z_5 = z_4 + \varepsilon$ to $\mathcal{N}$. Here it holds that $v(z_i) \geq 0$ for even i and $v(z_i) \leq 0$ for odd i.

The next lemma shows that a Tchebycheff system can also be characterized by $\widetilde{Z}$.

Lemma V.1.8 (Theorem I.4.2. in [38]). *Let $u_0, u_1, \ldots, u_n\colon [a,b] \to \mathbb{R}$ be continuous functions, $n \in \mathbb{N}$. $(u_i)_{i=0}^n$ constitutes a Tchebycheff system over $[a,b]$ if and only if $\widetilde{Z}(v) \leq n$ for every non-trivial u-polynomial v.*

Proof. First let $\widetilde{Z}(v) \leq n$ for any non-trivial u-polynomial v. Observe that $Z(v) \leq \widetilde{Z}(v)$, hence, Theorem V.1.6 implies that $(u_i)_{i=0}^n$ is a Tchebycheff system.

Now let $(u_i)_{i=0}^n$ be a Tchebycheff system and assume on the contrary that $\widetilde{Z}(v) \geq n+1$ for a non-trivial u-polynomial v. Denote by $\mathcal{N} = \{x_1, ..., x_k\}$, $a \leq x_1 < ... < x_k \leq b$, the set of distinct zeros of v in $[a, b]$. By Theorem V.1.6 it holds that $Z(v) \leq n$ and hence $k \leq n$. Thus, v has at least one nonnodal zero. Add the point $x_i + \varepsilon$ to $\mathcal{N}$ if x_i is a nonnodal zero, $1 \leq i \leq k$, and in addition, add the point $x_\ell - \varepsilon$ for the first nonnodal zero x_ℓ. Choose $\varepsilon > 0$ sufficiently small such that $x_i + \varepsilon < x_{i+1}$ for all $1 \leq i \leq k-1$, $x_\ell - \varepsilon > x_{\ell-1}$ if $\ell > 1$ and that the additional points are contained in $[a, b]$. Recall that a zero in a or b is a nodal zero and all nonnodal zeros are counted twice for $\widetilde{Z}$. Since for every nonnodal zero a point is added to $\mathcal{N}$ as well as an extra point for the smallest nonnodal zero, the enlarged set $\mathcal{N}$ consists of at least $n+2$ points. Now arrange the elements of $\mathcal{N}$ in increasing order and rename the first $n+2$ points as $z_0 < z_1 < ... < z_{n+1}$. Then the values $v(z_i)$ change their sign in the sense that $v(z_i) \geq 0$ for odd i and $v(z_i) \leq 0$ for even i or vice versa, see Figure V.1 for an example. Let

$$B = \begin{pmatrix} v(z_0) & v(z_1) & ... & v(z_{n+1}) \\ u_0(z_0) & u_0(z_1) & ... & u_0(z_{n+1}) \\ \vdots & \vdots & & \vdots \\ u_n(z_0) & u_n(z_1) & ... & u_n(z_{n+1}) \end{pmatrix}.$$

Then

$$0 = \det(B) = \sum_{j=0}^{n+1} (-1)^j v(z_j) \det(B_{1,j+1}), \tag{V.1.3}$$

where we expand the determinant along the first row and denote by $B_{1,j}$ the matrix arising from B by deleting the first row and jth column of the matrix B. Since $(u_i)_{i=0}^n$ is a Tchebycheff system the determinants $\det(B_{1,j+1})$ are either strictly positive for all $0 \leq j \leq n+1$ or strictly negative. Thus, $b_j := (-1)^j \det(B_{1,j+1})$, $0 \leq j \leq n+1$, strictly alternates in sign. Using (V.1.3) we obtain

$$\sum_{j=0}^{n+1} b_j v(z_j) = 0,$$

where either $b_j v(z_j) \geq 0$ for all $0 \leq j \leq n+1$ or every summand is less or equal to 0. Thus $v(z_j) = 0$ for all $0 \leq j \leq n+1$ and as a result v has at least $n+2$ different zeros. But this contradicts the fact that $Z(v) \leq n$ by Theorem V.1.6. To sum up, we have shown that $\widetilde{Z}(v) \leq n$. $\square$

In the sequel $(u_i)_{i=0}^n$, $n \in \mathbb{N}$, always denotes a Tchebycheff system over $[a, b]$.

V.1.2 u-Polynomials with a prescribed Zero Set

We construct non-negative u-polynomials with a prescribed set of zeros. For this purpose let $\mathcal{N} = \{x_1, ..., x_k\}$ with $a \leq x_1 < x_2 < ... < x_k \leq b$. To obtain a non-negative u-polynomial every zero in $\mathcal{N} \cap (a, b)$ has to be nonnodal. We assign a weight $\widetilde{w}(x_i)$ to each $x_i \in \mathcal{N}$ defined by

$$\widetilde{w}(x_i) = \begin{cases} 2, & x_i \in \mathcal{N} \cap (a, b), \\ 1, & x_i \in \mathcal{N} \cap \{a, b\}. \end{cases}$$

By Lemma V.1.8 a u-polynomial possessing exactly the zeros prescribed by the set $\mathcal{N}$ only exists if $\sum_{i=1}^k \widetilde{w}(x_i) \leq n$. Moreover, we have the following result.

Lemma V.1.9 (Theorem I.5.1.(a) in [38], based on Krein, [40])**.** *Let $(u_i)_{i=0}^n$, $n \in \mathbb{N}$, be a Tchebycheff system and let $\mathcal{N} = \{x_1, ..., x_k\} \subseteq [a, b]$, $k \in \mathbb{N}$, with $\sum_{i=1}^k \widetilde{w}(x_i) \leq n$.*

a) There exists a non-trivial non-negative u-polynomial w vanishing precisely at the points in $\mathcal{N}$ if n is even and either $a, b \in \mathcal{N}$ or $a, b \notin \mathcal{N}$ or if n is odd.

b) If n is even and exactly one of the endpoints a or b is contained in $\mathcal{N}$, then there exists a non-trivial non-negative u-polynomial w that vanishes at $\mathcal{N}$ and it may vanish at the other endpoint as well.

Proof. We first examine the case when $n = 2m + 1$ for some $m \geq 0$. Assume that $\mathcal{N} \subseteq (a, b)$. Then $\sum_{i=1}^{k} \widetilde{w}(x_i) = 2k \leq n = 2m + 1$ implies that $k \leq m$. If $k < m$, we extend the sequence $(x_i)_{i=1}^{k}$ by the points $a, x_1', ..., x_{m-k}'$ satisfying $x_k < x_1' < x_2' < ... < x_{m-k}' < b$ and we extend $\mathcal{N}$ by a if $k = m$. In the following we focus on the case $k < m$. Similar arguments apply for $k = m$. Let $\varepsilon > 0$ be sufficiently small such that the array $\big(z_i(\varepsilon)\big)_{i=1}^{2m+1}$ consisting of the points

$$a,\ x_1,\ x_1+\varepsilon,\ x_2,\ x_2+\varepsilon,\ \dots\ ,\ x_k,\ x_k+\varepsilon,\ x_1',\ x_1'+\varepsilon,\ x_2',\ x_2'+\varepsilon,\ \dots\ ,\ x_{m-k}',\ x_{m-k}'+\varepsilon$$

is in increasing order and $x_{m-k}' + \varepsilon < b$. If the determinant in (V.1.1) is strictly positive, we define the polynomial v_ε by

$$v_\varepsilon(x) = D\begin{pmatrix} 0 & 1 & \dots & 2m & 2m+1 \\ z_1(\varepsilon) & z_2(\varepsilon) & \dots & z_{2m+1}(\varepsilon) & x \end{pmatrix}. \tag{V.1.4}$$

Otherwise let

$$v_\varepsilon(x) = -D\begin{pmatrix} 0 & 1 & \dots & 2m & 2m+1 \\ z_1(\varepsilon) & z_2(\varepsilon) & \dots & z_{2m+1}(\varepsilon) & x \end{pmatrix}. \tag{V.1.5}$$

Since $(u_i)_{i=0}^{n}$ forms a Tchebycheff systems, v_ε vanishes precisely on the set $\{z_i(\varepsilon)\}_{i=1}^{2m+1}$. By Theorem V.1.8 we conclude that the zeros are nodal. Moreover, we have $v_\varepsilon(x) > 0$ for all $x > z_{2m+1}(\varepsilon)$. As a consequence, we find that

$$v_\varepsilon(x) > 0 \qquad \text{if } \ x \in \bigcup_{i=1}^{m} (z_{2i-1}, z_{2i}) \cup (z_{2m+1}, b]\,. \tag{V.1.6}$$

Expanding the determinant in (V.1.4) and (V.1.5), respectively, along the last column, we obtain

$$v_\varepsilon(x) = \sum_{i=0}^{n} a_i(\varepsilon) u_i(x)$$

with $\sum_{i=0}^{n} a_i^2(\varepsilon) > 0$. By multiplying with a suitable positive constant we can assume that $\sum_{i=0}^{n} a_i^2(\varepsilon) = 1$. Choose a sequence $(\varepsilon_k)_{k\in\mathbb{N}}$ with $\varepsilon_k \to 0$ as $k \to \infty$ such that $a(\varepsilon_k) := \big(a_i(\varepsilon_k)\big)_{i=0}^{n}$, $k \in \mathbb{N}$, converges to $a \in \mathbb{R}^{n+1}$ with $\sum_{i=0}^{n} a_i^2 = 1$ for $k \to \infty$. Denote by v the limiting polynomial $v(x) = \sum_{i=0}^{n} a_i u_i(x)$. Then v is non-negative by (V.1.6) and vanishes at $\mathcal{N}_v := \{a, x_1, x_2, ..., x_k, x_1', x_2', ..., x_{m-k}'\}$. Thus, the points x_i, $1 \leq i \leq k$ and x_j', $1 \leq j \leq m-k$, are nonnodal zeros and by Theorem V.1.8 the set of all distinct zeros of v is precisely given by $\mathcal{N}_v$. Similarly, we construct a non-negative u-polynomial $\widetilde{v}$ whose zero set is given by $\mathcal{N}_{\widetilde{v}} = \{x_1, x_2, ..., x_k, x_1'', x_2'', ..., x_{m-k}'', b\}$ with $x_k < x_1'' < .... < x_{m-k}'' < b$ and $\mathcal{N}_v \cap \mathcal{N}_{\widetilde{v}} = \mathcal{N}$. Then $w = v + \widetilde{v}$ forms a non-negative u-polynomial with prescribed zero set $\mathcal{N}$.

Now assume that $a, b \in \mathcal{N}$, i.e. $a = x_1 < x_2 < ... < x_{k-1} < x_k = b$. We conclude from $\sum_{i=1}^{n} \widetilde{w}(x_i) = 2k - 2 \leq n = 2m + 1$ that $k \leq m + 1$. In contrast to the previous case, we add to $\mathcal{N}$ the points $x_1' < x_2' < ... < x_{m-k+1}'$ satisfying $x_{k-1} < x_j' < x_k = b$ for all $1 \leq j \leq m - k + 1$ if $k < m + 1$. If $k = m + 1$, we do not change $\mathcal{N} = \{x_1, ..., x_k\}$. For $k < m + 1$ consider the array $\big(z_i(\varepsilon)\big)_{i=1}^{2m+1}$ given by

$$a = x_1,\ x_1+\varepsilon,\ x_2,\ x_2+\varepsilon,\ \dots\ ,\ x_{k-1},\ x_{k-1}+\varepsilon,\ x_1',\ x_1'+\varepsilon,\ \dots\ ,\ x_{m-k+1}',\ x_{m-k+1}'+\varepsilon,\ x_k = b$$

for $\varepsilon > 0$ sufficiently small such that $x_j + \varepsilon < x_{j+1}$ for $1 \leq j \leq k-2$, $x_{k-1} + \varepsilon < x_1'$, $x_i' + \varepsilon < x_{i+1}'$ for $1 \leq i \leq m - k$ and $x_{m-k+1}' + \varepsilon < x_k = b$. Here we define v_ε by (V.1.5) if the determinant in

(V.1.1) is strictly positive and by (V.1.4) if the determinant is strictly negative. The arguments from the case $a, b \notin \mathcal{N}$ imply the existence of a non-negative u-polynomial v with zero set $\{x_1 = a, x_2, ..., x_{k-1}, x_k = b, x_1', ..., x_{m-k+1}'\}$ and a non-negative function $\widetilde{v}$ vanishing at $\{x_1 = a, x_2, ..., x_k = b, x_1'', ..., x_{m-k+1}''\}$ with $x_{k-1} < x_1'' < ... < x_{m-k+1}'' < b$ and $\{x_i'\}_{i=1}^{m-k+1} \cap \{x_j''\}_{j=1}^{m-k+1} = \emptyset$. Therefore, $w := v + \widetilde{v}$ is a non-negative u-polynomial with zero set $\mathcal{N}$. If $k = m+1$, then we obtain w similarly.

If exactly one endpoint is contained in $\mathcal{N}$, then $k \leq m+1$. We add the points $\{x_j'\}_{j=1}^{m-k+1}$ with $x_k < x_1' < ... < x_{m-k+1}' < b$ if $a \in \mathcal{N}$ and $k < m+1$ and $x_{k-1} < x_1' < ... < x_{m-k+1}' < x_k = b$ if $b \in \mathcal{N}$ and $k < m+1$. If $k = m+1$, we add no point to $\mathcal{N}$. Similar arguments as in the case where no or both endpoints are contained in $\mathcal{N}$ show the existence of a non-negative u-polynomial w with zero set $\mathcal{N}$.

Finally, we turn to the case $n = 2m$, $m \in \mathbb{N}$. If $\mathcal{N} \subseteq (a, b)$, we proceed as in the case where n is odd and $a, b \notin \mathcal{N}$ with the only difference that we do not add the point a or b in the construction of v and $\widetilde{v}$, respectively. If $a, b \in \mathcal{N}$, we follow the reasoning for $n = 2m+1$ and $a, b \in \mathcal{N}$ but we do not add the point $x_1 + \varepsilon = a + \varepsilon$. If only one of the endpoints is contained in $\mathcal{N}$, then we also use the arguments from the case when n is odd and exactly one endpoint is contained in $\mathcal{N}$. We construct non-negative u-polynomials v and $\widetilde{v}$ vanishing at $A := \{x_1, x_2, ..., x_k, x_1', x_2', ..., x_{m-k}'\}$ and $\widetilde{A} := \{x_1, x_2, ..., x_k, x_1'', x_2'', ..., x_{m-k}''\}$, respectively, with $A \cap \widetilde{A} = \{x_1, ..., x_k\}$. Then $\widetilde{Z}(v), \widetilde{Z}(\widetilde{v}) \geq 2m - 1 = n - 1$. In this case we cannot apply Lemma V.1.6 to guarantee that the set of all distinct zeros of v and $\widetilde{v}$ is given by A and $\widetilde{A}$, respectively. Since v and $\widetilde{v}$ are non-negative, only the second endpoint can be an additional zero. Therefore, it is possible that $w = v + \widetilde{v}$ also vanishes at the second endpoint. $\square$

V.1.3 The Moment Space $\mathscr{M}_{n+1}$

In the following we introduce the moment space $\mathscr{M}_{n+1}$ associated to a Tchebycheff system $(u_i)_{i=0}^n$, $n \in \mathbb{N}$. The moment space contains the u-moments $\left(\int_a^b u_i(x) d\varsigma(x)\right)_{i=0}^n$ of a finite measure ν on $\big(\mathbb{R}, \mathcal{B}(\mathbb{R})\big)$ with Stieltjes measure function ς, i.e. $\nu\big((c,d]\big) = \varsigma(d) - \varsigma(c)$ for $a \leq c < d \leq b$ (see Theorem 1.1.2 in [24]). We set $\int_a^b g(x) d\varsigma(x) = \int_{[a,b]} g(x)\nu(dx)$ for a continuous and bounded function $g : \mathbb{R} \to \mathbb{R}$. Note that we do not assume that $\int_a^b 1 \, d\varsigma(x) = 1$.

Definition V.1.10. The *moment space* $\mathscr{M}_{n+1}$ associated to a Tchebycheff system $(u_i)_{i=0}^n$ is given by

$$\mathscr{M}_{n+1} = \left\{ \mathbf{c} = (c_0, ..., c_n) \in \mathbb{R}^{n+1} : \exists \varsigma \in \mathscr{S} \text{ such that } c_i = \int_a^b u_i(x) d\varsigma(x) \ \forall \, 0 \leq i \leq n \right\},$$

where $\mathscr{S}$ denotes the set of all non-decreasing right-continuous functions ς on $\mathbb{R}$ of bounded total variation. If $\mathbf{c} \in \mathscr{M}_{n+1}$ and $\varsigma \in \mathscr{S}$ such that $c_i = \int_a^b u_i(x) d\varsigma(x)$, $0 \leq i \leq n$, then we say that ς represents $\mathbf{c}$.

Now we show properties of the moment space.

Lemma V.1.11 (Theorem II.1.1. in [38])**.** *The moment space $\mathscr{M}_{n+1}$ is a closed convex cone.*

Proof. Let $\mathbf{c}, \mathbf{d} \in \mathscr{M}_{n+1}$ and $\varsigma^{\mathbf{c}}, \varsigma^{\mathbf{d}}$ such that for all $0 \leq i \leq n$

$$c_i = \int_a^b u_i(x) d\varsigma^{\mathbf{c}}(x) \qquad \text{and} \qquad d_i = \int_a^b u_i(x) d\varsigma^{\mathbf{d}}(x).$$

For $\lambda \in (0,1)$ and all $0 \leq i \leq n$ we obtain

$$\lambda c_i + (1-\lambda) d_i = \int_a^b \lambda u_i(x) d\varsigma^{\mathbf{c}}(x) + \int_a^b (1-\lambda) u_i(x) d\varsigma^{\mathbf{d}}(x) = \int_a^b u_i(x) d\tilde{\varsigma}(x),$$

where $\tilde{\varsigma}(x) = \lambda\varsigma^{\mathbf{c}}(x) + (1-\lambda)\varsigma^{\mathbf{d}}(x)$. Since $\tilde{\varsigma} \in \mathscr{S}$ we conclude that $\mathscr{M}_{n+1}$ is convex. For $\gamma > 0$ the function $\gamma\varsigma^{\mathbf{c}}$ is also a non-decreasing right-continuous function of bounded total variation. Thus,

$$\big(\gamma c_i\big)_{i=0}^n = \left(\int_a^b \gamma u_i(x) d\varsigma^{\mathbf{c}}(x)\right)_{i=0}^n = \left(\int_a^b u_i(x) d\big(\gamma\varsigma^{\mathbf{c}}(x)\big)\right)_{i=0}^n \in \mathscr{M}_{n+1}.$$

Hence, the moment space $\mathscr{M}_{n+1}$ is a cone.

To prove that $\mathscr{M}_{n+1}$ is closed, let $(\mathbf{c}^m)_{m\in\mathbb{N}}$ be a sequence in $\mathscr{M}_{n+1}$ which converges to $\mathbf{c} \in \mathbb{R}^{n+1}$ and denote by ς^m, $m \in \mathbb{N}$, a function in $\mathscr{S}$ representing $\mathbf{c}^m$. Let $v(x) = \sum_{i=0}^n a_i u_i(x)$ be a strictly positive u-polynomial. Choose $\mathcal{N} = \emptyset$ in Theorem V.1.9 to guarantee the existence of v. The boundedness of $(\mathbf{c}^m)_{m\in\mathbb{N}}$ implies that there exists $M \in (0,\infty)$ such that for all $m \in \mathbb{N}$ we have

$$\begin{aligned} M \geq \sum_{i=0}^n a_i c_i^m &= \sum_{i=0}^n \left(a_i \int_a^b u_i(x) d\varsigma^m(x)\right) = \int_a^b v(x) d\varsigma^m(x) \\ &\geq \left(\min_{a\leq r\leq b} v(r)\right) \int_a^b d\varsigma^m(x) = \left(\min_{a\leq r\leq b} v(r)\right) \big(\varsigma^m(b) - \varsigma^m(a-)\big), \end{aligned}$$

where $\varsigma^m(a-) = \lim_{x\uparrow a} \varsigma^m(x)$. Since the function v is strictly positive and continuous, its minimal value over $[a,b]$ is strictly positive. Hence, the differences $\varsigma^m(b) - \varsigma^m(a-)$, $m \in \mathbb{N}$, are uniformly bounded. Define

$$\hat{\varsigma}_m(x) = \begin{cases} \varsigma^m(a-), & x < a, \\ \varsigma^m(x), & x \in [a,b] \\ \varsigma^m(b), & x > b. \end{cases}$$

Then $(\hat{\varsigma}_m)_{m\in\mathbb{N}}$ is uniformly bounded and $\hat{\varsigma}_m$ represents $\mathbf{c}^m$, $m \in \mathbb{N}$.

Recall Helly's selection theorem (see e.g. Satz 4.16 in [25]) and the Helly-Bray theorem (see e.g. Satz 4.14 in [25]):

Let $(F_m)_{m\in\mathbb{N}}$ be a sequence in $\mathscr{S}$ which is uniformly bounded. Then there exist $F \in \mathscr{S}$ and a subsequence $(F_{m_k})_{k\in\mathbb{N}}$ such that for all continuity points x of F it holds that

$$\lim_{k\to\infty} F_{m_k}(x) = F(x).$$

Furthermore, for all $\mathcal{C}(\mathbb{R})$ with compact support we have

$$\lim_{k\to\infty} \int_{\mathbb{R}} g(x) dF_{m_k}(x) = \int_{\mathbb{R}} g(x) dF(x).$$

Hence, let $\varsigma \in \mathscr{S}$ such that there exists a subsequence $(\hat{\varsigma}_{m_k})_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} \hat{\varsigma}_{m_k}(x) = \varsigma(x)$ for all continuity points of ς and $\lim_{k\to\infty} \int_{\mathbb{R}} g(x) d\hat{\varsigma}_{m_k}(x) = \int_{\mathbb{R}} g(x) d\varsigma(x)$, where $g \in \mathcal{C}(\mathbb{R})$ with compact support. Since ς has at most countably many points of discontinuity and it is increasing, we conclude that ς is constant on $(-\infty, a)$ and $[b,\infty)$. Extend the functions u_i, $0 \leq i \leq n$, to continuous functions g_i on $\mathbb{R}$ with compact support K, in particular $g_i = u_i$ on $[a,b]$ and $g_i(x) = 0$ for all $x \notin K$. Then for all $k \in \mathbb{N}$ it holds that

$$\int_{\mathbb{R}} g_i(x) d\hat{\varsigma}_{m_k}(x) = \int_a^b u_i(x) d\hat{\varsigma}_{m_k}(x), \qquad \int_{\mathbb{R}} g_i(x) d\varsigma(x) = \int_a^b u_i(x) d\varsigma(x).$$

Therefore, we conclude that

$$c_i = \lim_{k\to\infty} c_i^{m_k} = \lim_{k\to\infty} \int_a^b u_i(x) d\hat{\varsigma}_{m_k}(x) = \lim_{k\to\infty} \int_{\mathbb{R}} g_i(x) d\hat{\varsigma}_{m_k}(x) = \int_{\mathbb{R}} g_i(x) d\varsigma(x) = \int_a^b u_i(x) d\varsigma(x)$$

for all $0 \leq i \leq n$. Hence, the moment space $\mathscr{M}_{n+1}$ is closed. $\square$

Remark V.1.12. The moment space $\mathscr{M}_{n+1}$ is not contained in an n-dimensional hyperplane. To see this, let $x_0, x_1, ..., x_n$ be $n+1$ distinct values in $[a, b]$. Since $\varsigma^j(x) = \mathbb{1}_{[x_j, b]}(x) \in \mathscr{S}$, $0 \le j \le n$, it holds that $\mathbf{c}^j = \big(u_0(x_j), u_1(x_j), ..., u_n(x_j)\big) \in \mathscr{M}_{n+1}$ and the Tchebycheff property implies that $(\mathbf{c}^0, ..., \mathbf{c}^n)$ are linearly independent.

In order to state a different characterization of the moment space we first introduce the smallest convex cone containing a curve $\big(u_0(x), ..., u_n(x)\big)$, $a \le x \le b$.

Definition V.1.13. Let C_{n+1} be the curve in $\mathbb{R}^{n+1}$ given by

$$C_{n+1} = \left\{\gamma(x) = \big(u_0(x), u_1(x), ..., u_n(x)\big) \colon a \le x \le b\right\}.$$

Denote by $\mathscr{C}(C_{n+1})$ the smallest convex cone containing C_{n+1}.

In Figure V.2 we provide an example for C_2 and the smallest convex cone $\mathscr{C}(C_2)$.

Lemma V.1.14. *The convex cone $\mathscr{C}(C_{n+1})$ is closed. Moreover, for every $\gamma = (\gamma_0, ..., \gamma_n) \in \mathscr{C}(C_{n+1})$ there exist $\lambda_j \ge 0$ and $a \le x_j \le b$, $1 \le j \le n+2$, such that*

$$\gamma_i = \sum_{j=1}^{n+2} \lambda_j u_i(x_j), \qquad 0 \le i \le n.$$

Proof. Let

$$\Gamma = \left\{\gamma = (\gamma_0, ..., \gamma_n) \in \mathbb{R}^{n+1} \colon \gamma_i = \sum_{j=1}^{n+2} \lambda_j u_i(x_j),\ 0 \le i \le n,\ \lambda_j \ge 0,\ x_j \in [a, b],\ 1 \le j \le n+2\right\}.$$

First observe that every $\gamma \in \Gamma$ is contained in the convex conical hull of C_{n+1}. Thus, $\Gamma \subseteq \mathscr{C}(C_{n+1})$. For the reverse inclusion recall Carathéodory's Theorem (see e.g. Theorem 17.1 in [57]): Every element in the convex hull of a set $A \subseteq \mathbb{R}^{n+1}$ can be written as a convex combination of at most $n+2$ points of A. Therefore, we conclude that $\Gamma = \mathscr{C}(C_{n+1})$.

In order to show that $\mathscr{C}(C_{n+1})$ is closed, let $(\gamma^m)_{m \in \mathbb{N}}$ be a sequence in $\mathscr{C}(C_{n+1})$ which converges to $\gamma = (\gamma_i)_{i=0}^n \in \mathbb{R}^{n+1}$ with respect to the Euclidean metric. For $m \in \mathbb{N}$ let $\lambda_j^m \ge 0$, $1 \le j \le n+2$, and $a \le x_1^m \le ... \le x_{n+2}^m \le b$ such that

$$\gamma_i^m = \sum_{j=1}^{n+2} \lambda_j^m u_i(x_j^m)$$

for all $0 \le i \le n$. Define the function $\varsigma^m \colon [a, b] \to \mathbb{R}$ by

$$\varsigma^m(x) = \sum_{j=1}^{n+2} \lambda_j^m \mathbb{1}_{[x_j^m, b]}(x).$$

Then it holds that $\varsigma^m \in \mathscr{S}$, $m \in \mathbb{N}$, and

$$\gamma_i^m = \int_a^b u_i(x) d\varsigma^m(x), \qquad 0 \le i \le n.$$

Now we proceed as in the proof of Lemma V.1.11 to conclude that there exist $\varsigma \in \mathscr{S}$ and a subsequence $(\varsigma^{m_k})_{k \in \mathbb{N}}$ with $\lim_{k\to\infty} \varsigma^{m_k}(x) = \varsigma(x)$ for all continuity points of ς and

$$\gamma_i = \lim_{k\to\infty} \gamma_i^{m_k} = \lim_{k\to\infty} \int_a^b u_i(x) d\varsigma^{m_k}(x) = \int_a^b u_i(x) d\varsigma(x) \tag{V.1.7}$$

for all $0 \le i \le n$. In addition, since ς^{m_k} is a non-negative step function with at most $n+2$ jumps, ς can be represented by

$$\varsigma(x) = \sum_{j=1}^{n+2} \lambda_j \mathbb{1}_{[x_j, b]}(x)$$

for some $\lambda_j \ge 0$ and $a \le x_j \le b$, $1 \le j \le n+2$. Therefore, $\gamma \in \Gamma$ and hence, $\Gamma = \mathscr{C}(C_{n+1})$ is closed. $\square$

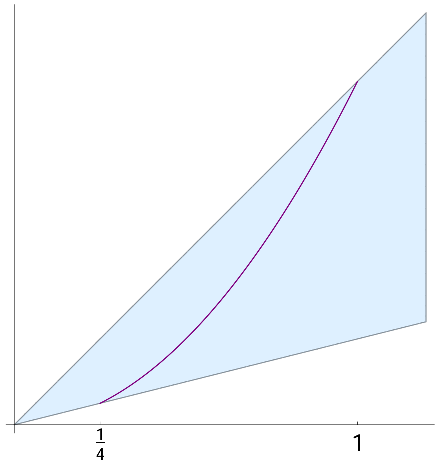


Figure V.2: For $u_0, u_1 \colon \left[\frac{1}{4}, 1\right] \to \mathbb{R}$ with $u_0(x) = x$ and $u_1(x) = x^2$ the violet line represents C_2 and the moment space $\mathscr{M}_2$ restricted to $[0, 1.2] \times [0, 1.2]$ is given by the blue area.

Next we show that the moment space coincides with $\mathscr{C}(C_{n+1})$.

Theorem V.1.15 (Theorem II.1.2. in [38])**.** *The moment space $\mathscr{M}_{n+1}$ is given by the convex conical hull of C_{n+1}. More precisely, $\mathscr{M}_{n+1} = \mathscr{C}\big(C_{n+1}\big)$.*

Proof. By Lemma V.1.14 every element $\gamma \in \mathscr{C}(C_{n+1})$ can be represented by

$$\gamma_i = \sum_{j=1}^{n+2} \lambda_j u_i(x_j) = \int_a^b u_i(x) d\varsigma(x),$$

where $\varsigma(x) := \sum_{j=1}^{n+2} \lambda_j \mathbb{1}_{[x_j, b]}(x) \in \mathscr{S}$, $\lambda_j \geq 0$ and $a \leq x_j \leq b$, $1 \leq j \leq n+2$. Hence, $\mathscr{C}(C_{n+1}) \subseteq \mathscr{M}_{n+1}$. Now let $\mathbf{c}^0 \in \mathscr{M}_{n+1}$ and assume that $\mathbf{c}^0 \notin \mathscr{C}(C_{n+1})$. Consider a hyperplane which strictly separates $\mathbf{c}^0$ from the closed convex set $\mathscr{C}(C_{n+1})$. More precisely, let $d \in \mathbb{R}$ and $a_0, a_1, ..., a_n \in \mathbb{R}$ with $\sum_{i=0}^n a_i^2 > 0$ such that

$$\sum_{i=0}^n a_i c_i^0 + d < 0 \quad \text{and} \quad \sum_{i=0}^n a_i \gamma_i + d \geq 0 \tag{V.1.8}$$

for all $\gamma = (\gamma_0, ..., \gamma_n) \in \mathscr{C}(C_{n+1})$. Since $\gamma = (0, ..., 0) \in \mathscr{C}(C_{n+1})$ we conclude that $d \geq 0$. Moreover, $\big(\lambda u_i(x)\big)_{i=0}^n \in \mathscr{C}(C_{n+1})$ for all $\lambda > 0$, hence for $a \leq x \leq b$ it holds that

$$\sum_{i=0}^n \lambda a_i u_i(x) + d \geq 0. \tag{V.1.9}$$

Let $\varsigma^0 \in \mathscr{S}$ with

$$c_i^0 = \int_a^b u_i(x) d\varsigma^0(x), \qquad 0 \leq i \leq n.$$

By (V.1.8) it follows that

$$-d > \sum_{i=0}^n a_i c_i^0 = \int_a^b \left(\sum_{i=0}^n a_i u_i(x)\right) d\varsigma^0(x). \tag{V.1.10}$$

On the other hand, integrating (V.1.9) with respect to ς^0 yields

$$0 \leq \int_a^b \left(\sum_{i=0}^n \lambda a_i u_i(x) + d\right) d\varsigma^0(x) = \lambda \left(\int_a^b \left(\sum_{i=0}^n a_i u_i(x)\right) d\varsigma^0(x) + \frac{d}{\lambda} \int_a^b d\varsigma^0(x)\right).$$

For $\lambda = \int_a^b d\varsigma^0(x) > 0$ we obtain a contradiction to (V.1.10). Here ς^0 is not constant, because otherwise $\mathbf{c}^0 = 0$, which is contained in $\mathscr{C}(C_{n+1})$. Therefore, $\mathscr{M}_{n+1} = \mathscr{C}(C_{n+1})$. $\square$

V.1.4 Boundary of $\mathscr{M}_{n+1}$

So far we have seen that every element in $\mathscr{M}_{n+1}$ can be written as a convex conical combination of at most $n+2$ points of $\mathscr{M}_{n+1}$. To characterize the boundary points of the moment space $\mathscr{M}_{n+1}$ we introduce the notion of an *index*.

Definition V.1.16. The *index* $\mathcal{I}(\mathbf{c})$ of a point $\mathbf{c} \in \mathscr{M}_{n+1}$ is the minimal number of points in C_{n+1} that are used in a convex conical representation of $\mathbf{c}$ with the convention that $\big(u_0(a), u_1(a), ..., u_n(a)\big)$ and $\big(u_0(b), u_1(b), ..., u_n(b)\big)$ are counted as half points whereas $\big(u_0(x), u_1(x), ..., u_n(x)\big)$ with $a < x < b$ is counted as 1.

Now we characterize the boundary $\partial\mathscr{M}_{n+1}$ of the moment space $\mathscr{M}_{n+1}$ in terms of the index.

Theorem V.1.17 (Theorem II.2.1. in [38]). *A point $\mathbf{c}^0 \in \mathscr{M}_{n+1}$, $\mathbf{c}^0 \neq 0$, is a boundary point of $\mathscr{M}_{n+1}$ if and only if $\mathcal{I}(\mathbf{c}^0) < \frac{n+1}{2}$. Moreover, for every boundary point $\mathbf{c}^0$ there exists a uniquely defined $p \leq \frac{n+2}{2}$ and unique $\lambda_j > 0$ and $a \leq x_j \leq b$, $1 \leq j \leq p$ such that*

$$c_i^0 = \sum_{j=1}^{p} \lambda_j u_i(x_j) \tag{V.1.11}$$

for all $0 \leq i \leq n$.

Proof. Let $\mathbf{c}^0 \in \mathscr{M}_{n+1}$, $\mathbf{c}^0 \neq 0$, with $\mathcal{I}(\mathbf{c}^0) < \frac{n+1}{2}$. We prove that $\mathbf{c}^0$ is a boundary point of $\mathscr{M}_{n+1}$. Let $\mathbf{c}^0$ be represented by

$$c_i^0 = \sum_{j=1}^{p} \lambda_j u_i(x_j),$$

where $p \leq \frac{n+2}{2}$, $\lambda_j > 0$ and $a \leq x_j \leq b$ for $1 \leq j \leq p$. Let $\mathcal{N} = \{x_1, ..., x_p\}$ and note that $\mathcal{I}(\mathbf{c}^0) < \frac{n+1}{2}$ implies that $p \leq \frac{n}{2}$ if $a, b \notin \mathcal{N}$, $p \leq \frac{n+2}{2}$ if $a, b \in \mathcal{N}$ and $p \leq \frac{n+1}{2}$ if exactly one of the endpoints is contained in $\mathcal{N}$. By Theorem V.1.9 there exists a non-negative u-polynomial $v^0(x) = \sum_{i=0}^{n} a_i u_i(x)$, $a_0, ..., a_n \in \mathbb{R}$, vanishing precisely at $\mathcal{N}$ or v^0 may vanishes at $\mathcal{N} \cup \{a, b\}$ if n is even and exactly one endpoint is contained in $\mathcal{N}$. Moreover, we have

$$\sum_{i=0}^{n} a_i c_i^0 = \sum_{i=0}^{n} a_i \sum_{j=1}^{p} \lambda_j u_i(x_j) = \sum_{j=1}^{p} \lambda_j \sum_{i=0}^{n} a_i u_i(x_j) = \sum_{j=0}^{p} \lambda_j v^0(x_j) = 0,$$

because the zero set of v^0 contains $\{x_1, ..., x_p\}$. For $\mathbf{c} \in \mathscr{M}_{n+1}$ and a function $\varsigma^{\mathbf{c}}$ representing $\mathbf{c}$ we obtain

$$\sum_{i=0}^{n} a_i c_i = \sum_{i=0}^{n} a_i \int_a^b u_i(x) d\varsigma^{\mathbf{c}}(x) = \int_a^b \left(\sum_{i=0}^{n} a_i u_i(x)\right) d\varsigma^{\mathbf{c}}(x) = \int_a^b v^0(x) d\varsigma^{\mathbf{c}}(x) \geq 0,$$

because v^0 is non-negative. Hence, $(a_i)_{i=0}^n$ determines a supporting hyperplane to $\mathscr{M}_{n+1}$ at $\mathbf{c}^0$, i.e.

$$\sum_{i=0}^{n} a_i c_i^0 = 0 \quad \text{and} \quad \sum_{i=0}^{n} a_i c_i \geq 0 \quad \forall\, \mathbf{c} \in \mathscr{M}_{n+1}.$$

Then we deduce from Lemma A.4.1 in Appendix A.4 that $\mathbf{c}^0$ is a boundary point of $\mathscr{M}_{n+1}$.

Let $\mathbf{c}^0 \in \partial\mathscr{M}_{n+1} \subseteq \mathscr{M}_{n+1}$, $\mathbf{c}^0 \neq 0$. We show that $\mathcal{I}(\mathbf{c}^0) < \frac{n+1}{2}$. Since $\mathscr{M}_{n+1}$ is convex, there exists a supporting hyperplane to $\mathscr{M}_{n+1}$ at $\mathbf{c}^0$, i.e. there exist $a = (a_i)_{i=0}^n \in \mathbb{R}^{n+1} \setminus \{0\}$ such that

$$\sum_{i=0}^{n} a_i c_i^0 + d = 0 \quad \text{and} \quad \sum_{i=0}^{n} a_i c_i + d \geq 0 \quad \text{for all } \mathbf{c} \in \mathscr{M}_{n+1}. \tag{V.1.12}$$

Since $\mathbf{c} = (0, 0, ..., 0) \in \mathscr{M}_{n+1}$ we obtain that $d \geq 0$ by (V.1.12). Assume that $d > 0$. For every $\lambda > 0$ the point $\lambda \mathbf{c}^0$ is contained in $\mathscr{M}_{n+1}$. Thus, (V.1.12) implies

$$0 \leq \sum_{i=0}^{n} \lambda a_i c_i^0 + d = \lambda \sum_{i=0}^{n} a_i c_i^0 + d = (1 - \lambda) d < 0$$

for all $\lambda > 1$. Hence, $d = 0$ and (V.1.12) simplifies to

$$\sum_{i=0}^{n} a_i c_i^0 = 0 \quad \text{ and } \quad \sum_{i=0}^{n} a_i c_i \geq 0 \quad \forall\, \mathbf{c} \in \mathscr{M}_{n+1}. \tag{V.1.13}$$

Now let $v^0(x) = \sum_{i=0}^{n} a_i u_i(x)$ and let $\varsigma^0 \in \mathscr{S}$ represent $\mathbf{c}^0$. Integrating v^0 with respect to ς^0 combined with (V.1.13) yields

$$\int_a^b v^0(x) d\varsigma^0(x) = \sum_{i=0}^{n} a_i \left(\int_a^b u_i(x) d\varsigma^0(x) \right) = \sum_{i=0}^{n} a_i c_i^0 = 0.$$

Moreover, for every $a \leq x \leq b$ the point $\mathbf{c}^x = \big(u_0(x), u_1(x), ..., u_n(x)\big)$ is contained in $\mathscr{M}_{n+1}$. Hence, (V.1.13) entails that

$$v^0(x) = \sum_{i=0}^{n} a_i u_i(x) = \sum_{i=0}^{n} a_i c_i^x \geq 0, \quad x \in [a, b].$$

As a consequence, the set of all points where ς^0 is strictly increasing is contained in the zero set $\mathcal{N}_{v^0}$ of v^0, which is a finite set by Theorem V.1.6. Hence, the function ς^0 is constant except at finitely many points $x_1, ..., x_p \in \mathcal{N}_{v^0}$ with $p \leq Z(v^0) \leq n$. This implies that there exist $\lambda_j > 0$, $1 \leq j \leq p$, such that $c_i^0 = \sum_{j=1}^{p} \lambda_j u_i(x_j)$, $0 \leq i \leq n$. The index of $\mathbf{c}^0$ satisfies $\mathcal{I}(\mathbf{c}^0) \leq p \leq n$. To determine $\mathcal{I}(\mathbf{c}^0)$ more accurately, note that all interior zeros of v^0, i.e. all zeros in (a, b), are nonnodal because $v^0 \geq 0$. Recall that $\widetilde{Z}(v) \leq n$ by Theorem V.1.8. Hence, we investigate the following three cases:

1. If $v^0(a), v^0(b) > 0$, then $\mathcal{I}(\mathbf{c}^0) = p \leq Z(v^0) = \frac{\widetilde{Z}(v^0)}{2} \leq \frac{n}{2}$.
2. If $v^0(a) = 0 = v^0(b)$, then the number of interior zeros is at most $\frac{n-2}{2}$ and, hence, $\mathcal{I}(\mathbf{c}^0) \leq \frac{1}{2} + \frac{n-2}{2} + \frac{1}{2} = \frac{n}{2}$. In addition, it holds that $p \leq \frac{n+2}{2}$.
3. Finally, if v^0 vanishes at one end point and is strictly positive at the other one, then there are at most $\frac{n-1}{2}$ interior zeros and the index satisfies $\mathcal{I}(\mathbf{c}^0) \leq \frac{1}{2} + \frac{n-1}{2} = \frac{n}{2}$ and hence, $p \leq \frac{n+1}{2}$.

To sum up, $\mathbf{c}^0$ is a boundary point of $\mathscr{M}_{n+1}$ if and only if $\mathcal{I}(\mathbf{c}^0) \leq \frac{n}{2}$.

Now we show that any representation of $\mathbf{c}^0$ of the type (V.1.11) is unique. We extend the set $\mathcal{N}_{v^0}$ of all roots of v^0 by additional points such that it contains exactly $n+1$ elements $z_0 < z_1 < ... < z_n$. Since the determinant $D \begin{pmatrix} 0 & 1 & \dots & n \\ z_0 & z_1 & \dots & z_n \end{pmatrix}$ is different from 0, the coefficients $\lambda_1, ..., \lambda_p$ in (V.1.11) are uniquely determined by $\mathbf{c}^0$. More precisely,

$$\mathbf{c}^0 = \begin{pmatrix} u_0(z_0) & u_0(z_1) & \dots & u_0(z_n) \\ u_1(z_0) & u_1(z_1) & \dots & u_1(z_n) \\ \vdots & \vdots & & \vdots \\ u_0(z_0) & u_0(z_1) & \dots & u_n(z_n) \end{pmatrix} \begin{pmatrix} \tilde{\lambda}_0 \\ \tilde{\lambda}_1 \\ \vdots \\ \tilde{\lambda}_n \end{pmatrix}$$

has a unique solution $(\tilde{\lambda}_0, ..., \tilde{\lambda}_n)$. Thus, we have shown that for a given supporting hyperplane to $\mathscr{M}_{n+1}$ at $\mathbf{c}^0$ there exists a unique representation of $\mathbf{c}^0$ given by (V.1.11). For fixed $x_1, ..., x_p$ the weights $\lambda_1, ..., \lambda_p$ are unique. Keep in mind that the points $x_1, ..., x_p$ arise from the chosen supporting hyperplane which is in general not unique. To verify that $x_1, ..., x_p$ are also unique, assume that there exists $b = (b_i)_{i=0}^n \in \mathbb{R}^{n+1}$, $b \neq 0$, $b \neq a$, such that

$$\sum_{i=0}^{n} b_i c_i^0 = 0 \quad \text{and} \quad \sum_{i=0}^{n} b_i c_i \geq 0 \quad \forall\, \mathbf{c} \in \mathscr{M}_{n+1}.$$

Let $w^0(x) = \sum_{i=0}^n b_i u_i(x)$. Using the same arguments as for v^0 we deduce that the set of all points where ς^0 is strictly increasing is contained in the zero set $\mathcal{N}_{w^0}$ of w^0. Note that $\widetilde{\mathcal{N}} := \mathcal{N}_{v^0} \cap \mathcal{N}_{w^0}$ is non-empty, because ς^0 represents $\mathbf{c}^0 \neq 0$ and hence, has at least one point of increase. Denote by $\ell \leq \min\{Z(v^0), Z(w^0)\} \leq n$ the number of elements in $\widetilde{\mathcal{N}}$. Since the weights $(\overline{\lambda}_j)_{j=1}^p$ in a representation $c_i^0 = \sum_{j=1}^p \overline{\lambda}_j u_i(z_j)$ with $\overline{\lambda}_j \geq 0$ and $z_j \in \widetilde{\mathcal{N}}$ are uniquely defined, it follows that for every boundary point $\mathbf{c}^0$ there exists precisely one representation. This representation is given by (V.1.11) with $p \leq \frac{n}{2}$ if $a, b \notin \widetilde{\mathcal{N}}$, $p \leq \frac{n+2}{2}$ if $a, b \in \widetilde{\mathcal{N}}$ and $p \leq \frac{n+2}{2}$ if exactly one endpoint is contained in $\widetilde{\mathcal{N}}$. □

V.1.5 Interior of $\mathscr{M}_{n+1}$

Next we focus on a characterization of the interior of $\mathscr{M}_{n+1}$ and introduce the concept of the *roots* and the *index of a representation.*

Definition V.1.18. Let $\mathbf{c} \in \mathscr{M}_{n+1}$ be given by

$$c_i = \sum_{j=1}^{p} \lambda_j u_i(x_j), \quad 0 \leq i \leq n, \tag{V.1.14}$$

where $\lambda_j > 0$, $a \leq x_j \leq b$, $1 \leq j \leq p$. We call the right-hand side of (V.1.14) a representation of $\mathbf{c}$ and $\{x_j\}_{j=1}^p$ the *roots of the representation* (V.1.14). Furthermore, we say that $\mathbf{c}$ *involves* $x_1, ..., x_p$ and we always assume that $x_1 < x_2 < ... < x_p$. The function $\varsigma^{\mathbf{c}}$ given by

$$\varsigma^{\mathbf{c}}(x) = \sum_{j=1}^{p} \lambda_j \mathbb{1}_{[x_j, b]}(x)$$

is called the *(associated) measure* of the representation (V.1.14). The *index* of a finite set $\mathcal{N} \subseteq [a, b]$ is defined as the number of elements in $\mathcal{N}$ with the convention that interior points are counted as one and the endpoints a and b as one half. The *index of the representation* (V.1.14) is the index of its set of roots in the representation and the *index of an associated measure* is also the index of its roots in the representation (V.1.14).

In the following we will emphasize whether we focus on the index of a point $\mathbf{c} \in \mathscr{M}_{n+1}$ (see Definition V.1.16) or the index of a representation as given in Definition V.1.18.

The following theorem shows that for any interior point of $\mathscr{M}_{n+1}$ and any $x^* \in [a, b]$ there exists a representation of index $\frac{n+1}{2}$ or $\frac{n+2}{2}$ such that x^* is a root of this representation. Recall that this is in contrast to the boundary points of $\mathscr{M}_{n+1}$ for which the representation is unique by Theorem V.1.17.

Theorem V.1.19 (Theorem II.3.1. in [38])**.** *Let $\mathbf{c}^0 \in int(\mathscr{M}_{n+1})$ and $x^* \in [a, b]$. Then there exists a representation of $\mathbf{c}^0$ of index $\frac{n+1}{2}$ or $\frac{n+2}{2}$ which involves the point x^*. More precisely, there exist a set $\{x_1, ..., x_p\}$ of index $\frac{n+1}{2}$ or $\frac{n+2}{2}$ with $x^* \in \{x_1, ..., x_p\}$ and $\lambda_j > 0, 1 \leq j \leq p$, such that*

$$c_i^0 = \sum_{j=1}^{p} \lambda_j u_i(x_j), \quad 0 \leq i \leq n.$$

Proof. Let $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$. Let $v = \sum_{i=0}^n a_i u_i$ be a strictly positive u-polynomial and set $\lambda = \sum_{i=0}^n a_i c_i^0$. Define $S_{n+1} = \{\mathbf{c} \in \mathscr{M}_{n+1} \colon \sum_{i=0}^n a_i c_i = \lambda\}$. Note that $\mathbf{c}^0$ is an interior point of S_{n+1}. For every $\mathbf{c} \in \mathscr{M}_{n+1}$, $\mathbf{c} \neq 0$, there exists a unique $\alpha > 0$ such that $\alpha\mathbf{c} \in S_{n+1}$. Moreover, S_{n+1} is bounded: Let $\mathbf{d} \in S_{n+1}$ and $\varsigma^{\mathbf{d}}$ such that $d_i = \int_a^b u_i(x)d\varsigma^{\mathbf{d}}(x)$ for all $0 \leq i \leq n$. Then

$$\lambda = \sum_{i=0}^n a_i d_i = \int_a^b v(x)d\varsigma^{\mathbf{d}}(x) \geq \left(\min_{a\leq r\leq b} v(r)\right)\int_a^b 1\, d\varsigma^{\mathbf{d}}(x).$$

Hence, for all $\varsigma \in \mathscr{S}$ which represent some $\mathbf{c} \in S_{n+1}$, it holds that $\int_a^b 1\, d\varsigma(x)$ is uniformly bounded. Moreover, the functions u_i, $0 \leq i \leq n$, are bounded on $[a,b]$. Hence, there exists $M \in (0,\infty)$ such that $c_i = \int_a^b u_i(x)d\varsigma(x) \in [-M, M]$, $0 \leq i \leq n$, for all $\mathbf{c} \in S_{n+1}$.

Choose $\alpha^* \in (0,\infty)$ such that $\mathbf{c}^* := \big(\alpha^* u_0(x^*), \alpha^* u_1(x^*), ..., \alpha^* u_n(x^*)\big) \in S_{n+1}$ and draw the line L from $\mathbf{c}^*$ through $\mathbf{c}^0$. Since S_{n+1} is bounded and for any $\mathbf{c} \in \partial\mathscr{M}_{n+1}$ there exists a unique $\gamma \in (0,\infty)$ such that $\gamma\mathbf{c} \in S_{n+1}$, we conclude that there exists an intersection point $\widetilde{\mathbf{c}} \neq \mathbf{c}^*$ of L and the boundary of $\mathscr{M}_{n+1}$. Thus,

$$\mathbf{c}^0 = \alpha\widetilde{\mathbf{c}} + (1-\alpha)\mathbf{c}^* \tag{V.1.15}$$

for some $\alpha \in (0,1)$. By construction, the index of $\mathbf{c}^*$ satisfies $\mathcal{I}(\mathbf{c}^*) \leq 1$. More precisely, $\mathcal{I}(\mathbf{c}^*) = \frac{1}{2}$ if $x^* \in \{a,b\}$ and $\mathcal{I}(\mathbf{c}^*) = 1$ if $x^* \in (a,b)$. Theorem V.1.17 implies that $\mathcal{I}(\widetilde{\mathbf{c}}) < \frac{n+1}{2}$ and $\mathcal{I}(\mathbf{c}^0) \geq \frac{n+1}{2}$. Moreover, $\mathcal{I}(\mathbf{c}^0) \leq \mathcal{I}(\widetilde{\mathbf{c}}) + \mathcal{I}(\mathbf{c}^*)$. Therefore, we conclude that $\mathcal{I}(\widetilde{\mathbf{c}})$ equals $\frac{n-1}{2}$ or $\frac{n}{2}$ if $x^* \in (a,b)$ and $\mathcal{I}(\widetilde{\mathbf{c}}) = \frac{n}{2}$ if $x^* \in \{a,b\}$. The representation of $\mathbf{c}^0$ given in (V.1.15) has index $\frac{n+1}{2}$ or $\frac{n+2}{2}$ and involves the point x^*. □

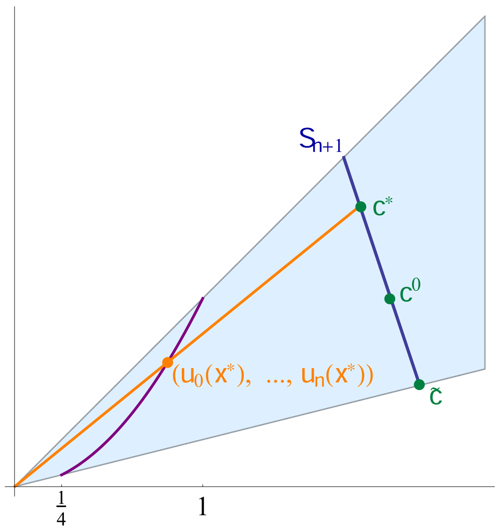


Figure V.3: The set S_{n+1} and the points $\mathbf{c}^0, \mathbf{c}^*$ and $\widetilde{\mathbf{c}}$ used in the proof of Theorem V.1.19 for $u_0(x) = x$ and $u_1(x) = x^2$, $x \in \left[\frac{1}{4}, 1\right]$.

Definition V.1.20. Let $\mathbf{c}^0$ be an interior point of $\mathscr{M}_{n+1}$. A representation for $\mathbf{c}^0$ of index $\frac{n+1}{2}$ is called *principal* and a representation of index $\leq \frac{n+2}{2}$ is called *canonical*. Furthermore, a canonical or principal representation is specified to be an *upper* (*lower*) *representation* if it involves the endpoint b (does not involve b).

Theorem V.1.19 shows that for any $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$ and $x^* \in [a,b]$ there exists a canonical representation involving x^* with positive weight. Now we show that for every $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$ there exist at least two principal representations, one is an upper representation whereas the second is a lower representation. First let $n = 2m$. Then we obtain two different principal representations by using the arguments from the proof of Theorem V.1.19 for $x^* = a$ as well as for $x^* = b$. Indeed, for $x^* = a$ the index of the point $\widetilde{\mathbf{c}}$, which is constructed in the proof of Theorem V.1.19, satisfies $\mathcal{I}(\widetilde{\mathbf{c}}) = \frac{n}{2} = m$ and a is not contained in any representation index $\frac{n}{2}$. Hence, the endpoint b cannot be involved in a representation of $\widetilde{\mathbf{c}}$ because otherwise the index is not an integer. To sum up, if $x^* \in \{a, b\}$, then the index of the representation (V.1.15) for $\mathbf{c}^0$ equals $\frac{n+1}{2}$ and involves exactly one endpoint a or b. Thus, we can specify two different principal representations for $\mathbf{c}^0$ with roots

$$a = x_1^* < x_2^* < ... < x_{m+1}^* < b \qquad \text{and} \qquad z_1^* < z_2^* < ... < z_m^* < z_{m+1}^* = b \tag{V.1.16}$$

for the lower and upper principal representation, respectively. Now consider the case $n = 2m + 1$. If we prescribe $x^* = a$, then by using the arguments from the proof of Theorem V.1.19 we obtain a principal representation of $\mathbf{c}^0$ which also involves the point b, because $\mathcal{I}(\widetilde{\mathbf{c}}) = m + \frac{1}{2}$. Hence, (V.1.15) is an upper principal representation for $\mathbf{c}^0$ with roots

$$a = z_1^* < z_2^* < ... < z_{m+1}^* < z_{m+2}^* = b. \tag{V.1.17}$$

To construct a lower principal representation for $\mathbf{c}^0$, consider a segment of the curve C_{n+1} given by $C_{n+1}(d) = \{\gamma(x) = (u_0(x), u_1(x), ..., u_n(x)) \colon d \leq x \leq b\}$, $d \in (a, b]$, and the convex conical hull $\mathscr{M}_{n+1}(d)$ of the segment $C_{n+1}(d)$. We claim that there exists $d' \in (a, b]$ such that $\mathbf{c}^0 \in \partial\mathscr{M}_{n+1}(d')$. Otherwise we can decompose the open interval (a, b) into two disjoint non-empty open sets A_1 and A_2 with $A_1 = \{d \in (a, b) \colon \mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1}(d))\}$ and $A_2 = \{d \in (a, b) \colon \mathbf{c}^0 \notin \mathscr{M}_{n+1}(d)\}$. But this contradicts the fact that (a, b) is connected. Therefore, we have $\mathbf{c}^0 \in \partial\mathscr{M}_{n+1}(d')$ for some $d' \in (a, b]$. By Theorem V.1.17 there exists a unique representation of $\mathbf{c}^0$ relative to $[d', b]$ whose roots are restricted to the interval $[d', b]$ and the representation has index $\mathcal{I}|_{[d',b]}(\mathbf{c}^0) \leq \frac{n}{2} = m + \frac{1}{2}$. On the other hand $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$ and hence, $\mathcal{I}(\mathbf{c}^0) \geq \frac{n+1}{2} = m + 1$. Therefore, we conclude that b is not involved in the representation of $\mathbf{c}^0$ but d'. This implies that neither a is involved in the lower representation. Relative to $[a, b]$, the representation of $\mathbf{c}^0$ has index $\frac{n+1}{2}$ and its roots are given by

$$a < x_1^* < x_2^* < ... < x_{m+1}^* < b. \tag{V.1.18}$$

So far we have proven the existence of at least two different principal representations for $\mathbf{c}^0$. In the following we show that these are the only principal representations and that their roots interlace, i.e.

$$\begin{aligned} a = x_1^* < z_1^* < x_2^* < z_2^* < ... < x_{m+1}^* < z_{m+1}^* = b, \qquad &\text{if } n = 2m, \\ a = z_1^* < x_1^* < z_2^* < x_2^* ... < z_{m+1}^* < x_{m+1}^* < z_{m+2}^* = b, \qquad &\text{if } n = 2m + 1. \end{aligned}$$

Lemma V.1.21 (Lemma II.3.1. in [38])**.** *Let $\mathbf{c}^0 \in int(\mathscr{M}_{n+1})$ and let ς, $\hat{\varsigma} \in \mathscr{S}$ be two different measures satisfying $\int_a^b u_i(x) d\varsigma(x) = c_i^0 = \int_a^b u_i(x) d\hat{\varsigma}(x)$, $0 \leq i \leq n$. Assume that ς is canonical, i.e. ς has index $\frac{n+1}{2}$ or $\frac{n+2}{2}$. Then for any pair of interior roots $x_j < x_k$ of ς there exists a point of increase of $\hat{\varsigma}$ in (x_j, x_k). If ς is principal, this remains true if $x_j = a$ or $x_k = b$.*

Proof. Let x_j and x_{j+1} be two consecutive roots of ς. Since ς is canonical, it has index $\frac{n+1}{2}$ or $\frac{n+2}{2}$. In the second case we require that x_j and x_{j+1} are interior roots (because otherwise we cannot guarantee that the function v used below exists) whereas we also allow for $x_j = a$ or $x_{j+1} = b$ if ς is principal. Assume that $\hat{\varsigma}$ concentrates no mass inside (x_j, x_{j+1}). Using similar arguments as in the proof of Theorem V.1.9 we construct a non-trivial u-polynomial v with $v(x) = \sum_{i=0}^n a_i u_i(x)$ satisfying

$$v(x) \begin{cases} \geq 0, & x \notin [x_j, x_{j+1}], \\ < 0, & x \in (x_j, x_{j+1}), \end{cases} \tag{V.1.19}$$

and v vanishes at the roots of ς and in (a, b) it has no other zeros. In contrast to the proof of Theorem V.1.9, we add slightly modified points to obtain a polynomial satisfying (V.1.19). Then

x_j and x_{j+1} are nodal zeros and we have

$$\begin{aligned}0 &= \sum_{i=0}^{n} a_i \left(\int_a^b u_i(x)d\hat{\varsigma}(x) - \int_a^b u_i(x)d\varsigma(x) \right)\\ &= \int_a^b v(x)d\hat{\varsigma}(x) - \int_a^b u(x)d\varsigma(x)\\ &= \int_a^b v(x)d\hat{\varsigma}(x).\end{aligned}$$

The last equality follows from the fact that v equals zero at the roots of ς, i.e. at the points where ς has positive mass. Since $\hat{\varsigma}$ has no mass inside the interval (x_j, x_{j+1}) by assumption and since the function v is non-negative on $[a, x_j] \cup [x_{j+1}, b]$ we conclude that

$$0 = \int_a^b v(x)d\hat{\varsigma}(x) = \int_a^{x_j} v(x)d\hat{\varsigma}(x) + \int_{x_{j+1}}^b v(x)d\hat{\varsigma}(x) \geq 0.$$

Therefore, $\hat{\varsigma}$ is constant on $\{x \in [a, b] \colon v(x) \neq 0\}$. This implies that $\hat{\varsigma}$ can only increase at the zero set $\mathcal{N}_v$ of v. Recall that $Z(v) \leq n$ by Theorem V.1.6. Thus, $\hat{\varsigma}$ can be represented with less than $n + 1$ roots, i.e.

$$\hat{\varsigma}(x) = \sum_{j=0}^{n-1} \lambda_j \mathbb{1}_{[z_j,b]}(x),$$

where $\lambda_j \geq 0$, $0 \leq j \leq n-1$, and $a \leq z_0 < z_1 < \dots < z_{n-1} \leq b$. This implies that $c_i^0 = \sum_{i=0}^{n-1} \lambda_j u_i(z_j)$ for all $0 \leq i \leq n$. On the other hand, we have $c_i^0 = \int_a^b u_i(x)d\varsigma(x) = \sum_{j=1}^{p} \beta_j u_i(x_j)$ with $\beta_j > 0$ for $1 \leq j \leq p$, where p has to be determined such that the index of $\{x_1, ..., x_p\}$ equals the index of ς. Note that $\{x_1, x_2, ..., x_p\} \subseteq \{z_0, z_1, ..., z_{n-1}\}$. Let $z_n \in [a, b] \backslash \{z_0, ..., z_{n-1}\}$. Then

$$\mathbf{c}^0 = \begin{pmatrix} u_0(z_0) & u_0(z_1) & \dots & u_0(z_n) \\ u_1(z_0) & u_1(z_1) & \dots & u_1(z_n) \\ \vdots & \vdots & & \vdots \\ u_n(z_0) & u_n(z_1) & \dots & u_n(z_n) \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_n \end{pmatrix}$$

has a unique solution and hence, ς and $\hat{\varsigma}$ coincide. As a result $\hat{\varsigma}$ has mass inside the interval (x_j, x_{j+1}). □

Corollary V.1.22 (Corollary II.3.1. in [38])**.** *For each $\mathbf{c}^0 \in int(\mathscr{M}_{n+1})$ there exist precisely two principal representations. The roots of these representations strictly interlace.*

Proof. We have already shown that there exist an upper and a lower principal representation for every $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$ whose roots interlace by Lemma V.1.21. We denote the measure associated to the lower principal representation in (V.1.16) and (V.1.18), respectively, by $\underline{\varsigma}$. Assume that there exists another principal representation with associated measure ς which differs from the principal representations given in (V.1.16) and (V.1.18), respectively. If $n = 2m + 1$ and ς has $m + 1$ interior roots $a < x_1' < \dots < x_{m+1}' < b$, these roots have to interlace with those in (V.1.18). Without loss of generality assume that $x_1' < x_1^*$. Using the proof of Theorem V.1.9 we construct a non-trivial u-polynomial v with nonnodal zeros at x_j' for $2 \leq j \leq m+1$ and a nodal zero at x_1'. Then it holds that $v(x) \geq 0$ for $x \geq x_1'$ and $v(x_j^*) > 0$ for $1 \leq j \leq m+1$. Thus, we obtain the following contradiction

$$0 = \int_a^b v(x)d\underline{\varsigma}(x) - \int_a^b v(x)d\varsigma(x) = \int_a^b v(x)d\underline{\varsigma}(x) > 0.$$

In all other cases the roots of ς involve at least one of the endpoints a or b. For odd n both endpoints are involved in the roots of an upper principal representation whereas for even n either the point a or b is involved in the roots of an lower and upper principal representation, respectively. Hence, the representations ς and the principal representation with roots given by (V.1.16) or (V.1.17) have at least one common root. This contradicts Lemma V.1.21. $\square$

We denote the measures associated to the unique upper and lower principal representation by $\overline{\varsigma}$ and $\underline{\varsigma}$, respectively.

V.1.6 A Measure Optimization Problem

In the following we extend the Tchebycheff system $(u_i)_{i=0}^n$ by a continuous function $f\colon [a,b] \to \mathbb{R}$ such that the extended system $(u_0, ..., u_n, f)$ still constitutes a Tchebycheff system. Denote by $\mathscr{M}_{n+2}$ the moment space associated to $u_0, ..., u_n, u_{n+1} = f$. For a fixed $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$ we aim at maximizing and minimizing $\int_a^b f(x) d\varsigma(x)$ over all $\varsigma \in \mathscr{S}$ representing $\mathbf{c}^0$.

Definition V.1.23. Let $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$.

1. We denote by $R(\mathbf{c}^0)$ the set of all measures $\varsigma \in \mathscr{S}$ representing $\mathbf{c}^0$. More precisely,

$$R(\mathbf{c}^0) = \left\{ \varsigma \in \mathscr{S} \colon c_i^0 = \int_a^b u_i(x) d\varsigma(x) \ \forall\, 0 \le i \le n \right\}.$$

2. The values of f integrated with respect to a measure $\varsigma \in R(\mathbf{c}^0)$ are given by

$$J(\mathbf{c}^0) = \left\{ \int_a^b f(x) d\varsigma(x) \colon \varsigma \in R(\mathbf{c}^0) \right\}.$$

We now show that for $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$ the set $J(\mathbf{c}^0)$ is a compact interval.

Lemma V.1.24. *Let $\mathbf{c}^0 \in int(\mathscr{M}_{n+1})$. Then there exist $\underline{\gamma}, \overline{\gamma} \in \mathbb{R}$, $\underline{\gamma} \le \overline{\gamma}$, such that $J(\mathbf{c}^0) = [\underline{\gamma}, \overline{\gamma}]$.*

Proof. Let $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$. First observe that $J(\mathbf{c}^0)$ is convex, because $R(\mathbf{c}^0)$ is convex. Moreover, $J(\mathbf{c}^0)$ is bounded: Let $v(x) = \sum_{i=0}^n a_i u_i(x)$, $a_0, ..., a_n \in \mathbb{R}$ be strictly positive. For every $\varsigma \in R(\mathbf{c}^0)$ it holds that $\int_a^b 1\, d\varsigma(x) \le \sum_{i=0}^n a_i c_i^0 / \min_{a \le r \le b} v(r)$. Since f is continuous on $[a,b]$ it follows that $\int_a^b f(x) d\varsigma(x)$ is uniformly bounded. As a consequence, $J(\mathbf{c}^0)$ is bounded. It remains to show that $J(\mathbf{c}^0)$ is closed. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $J(\mathbf{c}^0)$ such that $\gamma_n \to \gamma \in \mathbb{R}$ as $n \to \infty$. Denote by ς_n the functions in $R(\mathbf{c}^0)$ with $\gamma_n = \int_a^b f(x) d\varsigma_n(x)$. Similar to the proof of Lemma V.1.11 we conclude that there exist a subsequence $(\varsigma_{n_k})_{k\in\mathbb{N}}$ and $\varsigma \in \mathscr{S}$ such that

$$\int_a^b u_i(x) d\varsigma(x) = \lim_{k\to\infty} \int_a^b u_i(x) d\varsigma_{n_k}(x) = c_i^0,$$

$$\int_a^b f(x) d\varsigma(x) = \int_a^b f(x) d\varsigma_{n_k}(x) = \lim_{k\to\infty} \gamma_{n_k} = \gamma.$$

Hence, $\varsigma \in R(\mathbf{c}^0)$ and thus, $\gamma \in J(\mathbf{c}^0)$. To summarize, $J(\mathbf{c}^0) = [\underline{\gamma}, \overline{\gamma}]$ for some $\underline{\gamma}, \overline{\gamma} \in \mathbb{R}$, $\underline{\gamma} \le \overline{\gamma}$. $\square$

Observe that the points $\overline{\mathbf{c}} = \left(c_0^0, c_1^0, ..., c_n^0, \overline{\gamma}\right)$ and $\underline{\mathbf{c}} = \left(c_0^0, c_1^0, ..., c_n^0, \underline{\gamma}\right)$ are boundary points of $\mathscr{M}_{n+2}$. Hence, for each of these boundary points there exists a unique representation of index $\mathcal{I}(\overline{\mathbf{c}}), \mathcal{I}(\underline{\mathbf{c}}) \le \frac{n+1}{2}$ by Theorem V.1.17. Since $\mathbf{c}^0 \in \operatorname{int}(\mathscr{M}_{n+1})$ the index of these representations equals $\frac{n+1}{2}$. There exists exactly two different representations $\overline{\varsigma}$ and $\underline{\varsigma}$ of index $\frac{n+1}{2}$ for $\mathbf{c}^0$ by Corollary V.1.22. Now we show that $\int_a^b f(x) d\overline{\varsigma}(x) \ne \int_a^b f(x) d\underline{\varsigma}(x)$, which implies that $J(\mathbf{c}^0)$ does not degenerate to a point. Since the representations of $\overline{\varsigma}$ and $\underline{\varsigma}$ have index $\frac{n+1}{2}$ and we have seen that the endpoints a and b are included in the common roots of the representations, there are n interior roots. Thus, we obtain

$$\underline{\varsigma}(x) = \sum_{k=0}^{p} \alpha_k \mathbb{1}_{[x_k^*, b]}(x), \qquad \overline{\varsigma}(x) = \sum_{\ell=0}^{n-p} \beta_\ell \mathbb{1}_{[z_\ell^*, b]}(x),$$

where $\alpha_k, \beta_\ell > 0$, $a \le x_k^* \le b$, $a \le z_\ell^* \le b$ for $0 \le k \le p$, $0 \le \ell \le n-p$, with p such that the index of the representation $\overline{\varsigma}$ equals $\frac{n+1}{2}$. Furthermore, we assume that the roots $\{x_k^*\}_{k=0}^p$ and $\{z_\ell^*\}_{\ell=0}^{n-p}$ are strictly increasing. By Lemma V.1.21 the roots $\{x_k^*\}_{k=0}^p$ and $\{z_\ell^*\}_{\ell=0}^{n-p}$ strictly interlace. In addition, the following equality holds true

$$\sum_{k=0}^{p} \alpha_k u_i(x_k^*) = \int_a^b u_i(x) d\underline{\varsigma}(x) = c_i^0 = \int_a^b u_i(x) d\overline{\varsigma}(x) = \sum_{\ell=0}^{n-p} \beta_\ell u_i(z_\ell^*) \tag{V.1.20}$$

for all $0 \le i \le n$ and we have

$$\int_a^b f(x) d\underline{\varsigma}(x) = \sum_{k=0}^{p} \alpha_k f(x_k^*), \qquad \int_a^b f(x) d\overline{\varsigma}(x) = \sum_{\ell=0}^{n-p} \beta_\ell f(z_\ell^*).$$

Denote by $\{r_m\}_{m=0}^{n+1}$, $a = r_0 < r_1 < ... < r_n < r_{n+1} = b$ the ordered set of the common roots of $\overline{\varsigma}$ and $\underline{\varsigma}$ and let $\{\rho_m\}_{m=0}^{n+1}$ be the corresponding coefficients, i.e.

$$\rho_m = \begin{cases} \alpha_k, & \text{if } \ r_m = x_k^* \ \text{ for some } 0 \le k \le p, \\ \beta_\ell, & \text{if } \ r_m = z_\ell^* \ \text{ for some } 0 \le \ell \le n-p. \end{cases}$$

Now assume that

$$\sum_{k=0}^{p} \alpha_k f(x_k^*) = \int_a^b f(x) d\underline{\varsigma}(x) = \int_a^b f(x) d\overline{\varsigma}(x) = \sum_{\ell=0}^{n-p} \beta_\ell f(z_\ell^*). \tag{V.1.21}$$

Since the extended system $(u_0, ..., u_n, f)$ is a Tchebycheff system, it follows by (V.1.20) and (V.1.21) that

$$0 < \left(\prod_{m=0}^{n+1} \rho_m\right) D\begin{pmatrix} 0 & 1 & ... & n+1 \\ r_0 & r_1 & ... & r_{n+1} \end{pmatrix} = \det \begin{pmatrix} \rho_0 u_0(r_0) & \rho_1 u_0(r_1) & ... & \rho_{n+1} u_0(r_{n+1}) \\ \rho_0 u_1(r_0) & \rho_1 u_1(r_1) & ... & \rho_{n+1} u_1(r_{n+1}) \\ \vdots & \vdots & & \vdots \\ \rho_0 u_n(r_0) & \rho_1 u_n(r_1) & ... & \rho_{n+1} u_n(r_{n+1}) \\ \rho_0 f(r_0) & \rho_1 f(r_1) & ... & \rho_{n+1} f(r_{n+1}) \end{pmatrix}$$

$$= \det \begin{pmatrix} \sum_{m=1}^{n+1} \widetilde{\rho}_m u_0(r_m) & \rho_1 u_0(r_1) & ... & \rho_{n+1} u_0(r_{n+1}) \\ \sum_{m=1}^{n+1} \widetilde{\rho}_m u_1(r_m) & \rho_1 u_1(r_1) & ... & \rho_{n+1} u_1(r_{n+1}) \\ \vdots & \vdots & & \vdots \\ \sum_{m=1}^{n+1} \widetilde{\rho}_m u_n(r_m) & \rho_1 u_n(r_1) & ... & \rho_{n+1} u_n(r_{n+1}) \\ \sum_{m=1}^{n+1} \widetilde{\rho}_m f(r_m) & \rho_1 f(r_1) & ... & \rho_{n+1} f(r_{n+1}) \end{pmatrix} = 0,$$

where we set

$$\widetilde{\rho}_m = \begin{cases} \alpha_k, & \text{if } \ r_m = x_k^*, \ \ 0 \le k \le p, \\ -\beta_\ell, & \text{if } \ r_m = z_\ell^*, \ \ 0 \le \ell \le n-p, \end{cases}$$

if $\rho_0 = \alpha_0$ and

$$\widetilde{\rho}_m = \begin{cases} -\alpha_k, & \text{if } \ r_m = x_k^*, \ \ 0 \le k \le p, \\ \beta_\ell, & \text{if } \ r_m = z_\ell^*, \ \ 0 \le \ell \le n-p, \end{cases}$$

if $\rho_0 = \beta_0$. To summarize, $J(\mathbf{c}^0)$ is not degenerated and $\overline{\mathbf{c}} \neq \underline{\mathbf{c}}$. Now we show that

$$\sup_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) = \overline{\gamma} = \int_a^b f(x) d\overline{\varsigma}(x),$$
$$\inf_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) = \underline{\gamma} = \int_a^b f(x) d\underline{\varsigma}(x)$$

if the determinants $D\begin{pmatrix} 0 & 1 & \dots & n \\ x_0 & x_1 & \dots & x_n \end{pmatrix}$ and $D\begin{pmatrix} 0 & 1 & \dots & n+1 \\ z_0 & z_1 & \dots & z_{n+1} \end{pmatrix}$ have the same sign for $a \leq x_0 < \dots < x_m \leq b$, $a \leq z_0 < \dots < z_n < z_{n+1} \leq b$ and

$$\sup_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) = \int_a^b f(x) d\underline{\varsigma}(x), \qquad \inf_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) = \int_a^b f(x) d\overline{\varsigma}(x),$$

otherwise. Let $\{(z_\ell^*, \beta_\ell)\}_{\ell=0}^p$ and $\{(x_k^*, \alpha_k)\}_{k=0}^q$ be the roots and corresponding weights involved in the unique representation of $\overline{\mathbf{c}}$ and $\underline{\mathbf{c}}$, respectively. Recall that these representations are principal for $\mathbf{c}^0$ and that the roots $\{x_k^*\}_{k=0}^q$ and $\{z_\ell^*\}_{\ell=0}^p$ strictly interlace by Corollary V.1.22. Moreover, the set of common roots $\{r_m^*\}_{m=0}^{n+1} = \{x_k^*\}_{k=0}^q \cup \{z_\ell^*\}_{\ell=0}^p$ contains $n+2$ distinct points. We assume that $a = r_0^* < r_1^* < \dots < r_n^* < r_{n+1}^* = b$. Similar to (V.1.20) we obtain for all $0 \leq i \leq n$ that

$$\begin{aligned} 0 &= \sum_{\ell=0}^p \beta_\ell u_i(z_\ell^*) - \sum_{k=0}^q \alpha_k u_i(x_k^*), \\ 0 < \overline{\gamma} - \underline{\gamma} &= \sum_{\ell=0}^p \beta_\ell f(z_\ell^*) - \sum_{k=0}^q \alpha_k f(x_k^*). \end{aligned} \tag{V.1.22}$$

Let $\{\rho_m\}_{m=0}^{n+1}$ be the corresponding weights for $\{r_m^*\}_{m=0}^{n+1}$, where $\rho_m = \beta_\ell$ if $r_m = z_\ell^*$ for some $0 \leq \ell \leq p$ and $\rho_m = -\alpha_k$ if $r_m = x_k^*$ for some $0 \leq k \leq q$. Thus, (V.1.22) can be rewritten as

$$\begin{pmatrix} u_0(r_0^*) & u_0(r_1^*) & \dots & u_0(r_{n+1}^*) \\ u_1(r_0^*) & u_1(r_1^*) & \dots & u_1(r_{n+1}^*) \\ \vdots & \vdots & & \vdots \\ u_n(r_0^*) & u_n(r_1^*) & \dots & u_n(r_{n+1}^*) \\ f(r_0^*) & f(r_1^*) & \dots & f(r_{n+1}^*) \end{pmatrix} \begin{pmatrix} \rho_0 \\ \rho_1 \\ \vdots \\ \rho_n \\ \rho_{n+1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \overline{\gamma} - \underline{\gamma} \end{pmatrix}.$$

Cramer's Rule implies that

$$\rho_{n+1} = \frac{\det \begin{pmatrix} u_0(r_0^*) & u_0(r_1^*) & \dots & u_0(r_n^*) & 0 \\ u_1(r_0^*) & u_1(r_1^*) & \dots & u_1(r_n^*) & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ u_n(r_0^*) & u_n(r_1^*) & \dots & u_n(r_n^*) & 0 \\ f(r_0^*) & f(r_1^*) & \dots & f(r_n^*) & \overline{\gamma} - \underline{\gamma} \end{pmatrix}}{D\begin{pmatrix} 0 & 1 & \dots & n & n+1 \\ r_0^* & r_1^* & \dots & r_n^* & r_{n+1}^* \end{pmatrix}} = (\overline{\gamma} - \underline{\gamma}) \frac{D\begin{pmatrix} 0 & 1 & \dots & n \\ r_0^* & r_1^* & \dots & r_n^* \end{pmatrix}}{D\begin{pmatrix} 0 & 1 & \dots & n & n+1 \\ r_0^* & r_1^* & \dots & r_n^* & r_{n+1}^* \end{pmatrix}},$$

which is strictly positive if the sign of determinants coincide and strictly negative otherwise. In the first case, this implies that $\rho_{n+1} = \beta_p$ and hence $z_p^* = b$ whereas we find that $x_q^* = b$ in the second case.

To summarize, we have shown the following Theorem.

Theorem V.1.25 (based on Theorem III.1.1. in [38]). *Let $u_0, ..., u_n, f\colon [a,b] \to \mathbb{R}$ be continuous functions and assume that $(u_i)_{i=0}^n$ and $(u_0, ..., u_n, f)$, are Tchebycheff systems. Let $\mathbf{c}^0 \in int(\mathscr{M}_{n+1})$ and denote by $\overline{\varsigma}$ and $\underline{\varsigma}$ the measures corresponding to the upper and lower principal representation of $\mathbf{c}^0$, respectively.*

1. *If the determinants $D\begin{pmatrix} 0 & 1 & ... & n \\ x_0 & x_1 & ... & x_n \end{pmatrix}$ and $D\begin{pmatrix} 0 & 1 & ... & n+1 \\ z_0 & z_1 & ... & z_{n+1} \end{pmatrix}$ are either both strictly positive or strictly negative for all $a \le x_0 < x_1 < ... < x_n \le b$ and $a \le z_0 < z_1 < ... < z_{n+1} \le b$, then*

$$\begin{aligned} \sup_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) &= \int_a^b f(x) d\overline{\varsigma}(x), \\ \inf_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) &= \int_a^b f(x) d\underline{\varsigma}(x). \end{aligned} \tag{V.1.23}$$

   Moreover, $\overline{\varsigma}$ and $\underline{\varsigma}$ are the unique optimal measures.

2. *If $D\begin{pmatrix} 0 & 1 & ... & n \\ x_0 & x_1 & ... & x_n \end{pmatrix} \cdot D\begin{pmatrix} 0 & 1 & ... & n+1 \\ z_0 & z_1 & ... & z_{n+1} \end{pmatrix} < 0$ with $a \le x_0 < x_1 < ... < x_n \le b$ and $a \le z_0 < z_1 < ... < z_{n+1} \le b$, then*

$$\begin{aligned} \sup_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) &= \int_a^b f(x) d\underline{\varsigma}(x), \\ \inf_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) &= \int_a^b f(x) d\overline{\varsigma}(x). \end{aligned} \tag{V.1.24}$$

   The maximum and minimum is uniquely attained by $\underline{\varsigma}$ and $\overline{\varsigma}$, respectively.

Remark V.1.26. It is remarkable that if $(u_i)_{i=0}^n$ and $(u_0, ..., u_n, f)$ are Tchebycheff systems, the maximizing and minimizing measures in (V.1.23) and (V.1.24), respectively, are independent of the function f.

Remark V.1.27. If $\mathbf{c}^0 \in \partial\mathscr{M}_{n+1}$, then it admits a unique representation by Theorem V.1.17. Thus, the set $R(\mathbf{c}^0)$ consists of precisely one element $\hat{\varsigma}$. In particular,

$$\sup_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) = \inf_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x) d\varsigma(x) = \int_a^b f(x) d\hat{\varsigma}(x).$$

V.2 Optimal Stopping Problems and Tchebycheff Systems

We consider a one-dimensional regular continuous strong Markov process $(Y_t)_{t\in[0,\infty)}$ as defined in Chapter III and assume that all conditions of Chapter III are satisfied. The interior of the state space J is given by (a,b), $-\infty \le a < b \le \infty$. We consider the optimal stopping problem

$$V(T,y) = \sup_{\tau \in \mathcal{S}(T,y)} \mathbb{E}^y[f(Y_\tau)], \tag{V.2.1}$$

where $\mathcal{S}(T,y)$ denotes the set of all $(\mathcal{F}_t)$-stopping times τ with $\mathbb{E}^y[\tau] \le T$ and $f\colon J \to \mathbb{R}$ is continuous. Theorem III.2.2 ensures that the stopping problem (V.2.1) is equivalent to the measure optimization problem

$$V(T,y) = \sup_{\mu \in \mathcal{A}(T,y)} \int_{\mathbb{R}} f(x)\mu(dx). \tag{V.2.2}$$

We derive sufficient conditions for f guaranteeing that a reduction in (V.2.2) to the set $\mathcal{A}_2(T,y)$ is possible, where $\mathcal{A}_2(T,y)$ consists of all measures $\mu \in \mathcal{A}(T,y)$ that are weighted sums of at most 2 Dirac measures.

Firstly, we focus on processes with state space $J = [a,b]$, $-\infty < a < b < \infty$. In particular, Y is a martingale and $\mathcal{A}(T,y) = \left\{\mu \in \mathcal{M}^1 \colon \int_{\mathbb{R}} x\,\mu(dx) = y, \int_{\mathbb{R}} q_y(x)\mu(dx) \le T\right\}$. We show that a reduction in (V.2.2) to the set $\mathcal{A}_2(T,y)$ is possible if the functions $u_0(x) = 1$, $u_1(x) = x$, $u_2(x) = q_y(x)$ and f form a Tchebycheff system over $[a,b]$. In particular, we can restrict the stopping times in (V.2.1) to those stopping times $\tau \in \mathcal{S}(T,y)$ such that the law of Y_τ is a weighted sum of at most 2 Dirac measures, i.e. τ is the first exit time of an interval.

For this purpose, let

$$\mathcal{A}_=(S,y) = \left\{\mu \in \mathcal{M}^1\big([a,b]\big) \colon \int_{\mathbb{R}} x\,\mu(dx) = y, \int_{\mathbb{R}} q_y(x)\mu(dx) = S\right\},$$

where $y \in (a,b)$ and $S \in [0,\infty)$. Define the auxiliary measure optimization problem

$$\widetilde{U}(S,y) = \sup_{\mu \in \mathcal{A}_=(S,y)} \int_{\mathbb{R}} f(x)\mu(dx) \tag{V.2.3}$$

with the convention that $\sup\emptyset = -\infty$. Observe that $\mathcal{A}(T,y) = \bigcup_{S\in[0,T]} \mathcal{A}_=(S,y)$, because all measures in $\mathcal{A}(T,y)$ are centered around y. Thus,

$$V(T,y) = \sup_{S\in[0,T]} \widetilde{U}(S,y).$$

Hence, in order to show that we can confine to the set $\mathcal{A}_2(T,y)$ in (V.2.2), it is sufficient to show that for all $S \in [0,T]$ the supremum in (V.2.3) is attained by a measure in $\mathcal{A}_=(S,y)$ which is a weighted sum of at most 2 Dirac measures.

Remark V.2.1.

a) The set $\mathcal{A}_=(S,y)$ coincides with the set of all probability measures that can be embedded into Y under $\mathbb{P}^y$ with a minimal stopping time τ satisfying $\mathbb{E}^y[\tau] = S$.

b) It holds that $\mathcal{A}_=(S,y) = \emptyset$ for all $S > \frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b) = \mathbb{E}^y[\tau_{a,b}]$. In particular,

$$V(T,y) = V\left(\frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b), y\right), \qquad T \ge \frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b).$$

To see this, let $\mu \in \mathcal{A}_=(S,y)$ for some $S \in [0,\infty)$ and let τ be a minimal stopping time that embeds μ into Y under $\mathbb{P}^y$. Using the convexity of q_y and Theorem 2.4 in [32] it follows that

$$\int_{\mathbb{R}} q_y(x)\mu(dx) = \mathbb{E}^y[q_y(Y_\tau)] \le \mathbb{E}^y\left[\frac{b-Y_\tau}{b-a}q_y(a) + \frac{Y_\tau - a}{b-a}q_y(b)\right] = \frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b),$$

because $\mathbb{E}^y[Y_\tau] = \int_{\mathbb{R}} x\,\mu(dx) = y$.

V.2.1 Reduction to weighted Sums of 2 Dirac Measures

We show that the functions $u_0(x) = 1$, $u_1(x) = x$ and $u_2(x) = q_y(x)$, $x \in [a,b]$, $y \in (a,b)$, constitute a Tchebycheff system over $[a,b]$. If the extended system (u_0,u_1,u_2,f) also forms a Tchebycheff system, then the results of Section V.1 allow us to determine the value of the measure optimization problem (V.2.3). Moreover, if $\mathbf{c}^0 = (1,y,T) \in \operatorname{int}(\mathscr{M}_3)$, then there exists a unique maximizer for (V.2.3) which does not depend on the function f.

Lemma V.2.2. *Let $y \in (a,b)$. The functions $u_0(x) = 1$, $u_1(x) = x$ and $u_2(x) = q_y(x)$ form a Tchebycheff system over $[a,b]$.*

Proof. Let $y \in (a,b)$. Observe that $q_y(x) < \infty$ for all $x \in [a,b]$ by Lemma III.1.1. In particular, q_y is continuous on $[a,b]$. Let $a \le x_0 < x_1 < x_2 \le b$. From the proof of Theorem III.2.6 for measurable payoff functions (see page 63) we conclude that the vectors $\left\{\left(1, x_j, q_y(x_j)\right)\right\}_{j=0}^{2}$ are linearly independent. Thus, the determinant $D\begin{pmatrix} 0 & 1 & 2 \\ x_0 & x_1 & x_2\end{pmatrix}$ in (V.1.1) does not vanish. Due to the continuity of the functions u_0, u_1 and u_2, the determinant in (V.1.1) is either strictly positive for all $a \le x_0 < x_1 < x_2 \le b$ or strictly negative. □

Notice that (III.2.2) implies that for $y, z \in (a,b)$

$$q_z(x) = q_y(x) + c_1 x + c_2, \quad t \in [a,b],$$

where $c_1 = -\frac{1}{2}\left(\frac{\partial^+ q_y}{\partial x}(z) + \frac{\partial^- q_y}{\partial x}(z)\right)$ and $c_2 = -c_1 z - q_y(z)$. Hence, if (u_0, u_1, q_y, f) is a Tchebycheff systems for some $y \in (a,b)$, then (u_0, u_1, q_z, f) constitutes a Tchebycheff system for all $z \in (a,b)$.

Now we state a sufficient condition for the function q_y and the payoff function f such that the extended system (u_0, u_1, u_2, f) is a Tchebycheff system.

Lemma V.2.3. *Let $-\infty < a < b < \infty$. Let $\eta\colon [a,b] \to \mathbb{R}\backslash\{0\}$ be measurable and such that η^2 is continuous. For $y \in (a,b)$ define $q_y(x) = \int_y^x \int_y^w \frac{2}{\eta^2(u)} du\, dw$, $x \in [a,b]$, and $q_y(x) = \infty$, $x \notin [a,b]$. Let $f\colon [a,b] \to \mathbb{R}$ be twice differentiable on (a,b) with continuous first derivative such that $\eta^2 f''$ is injective on (a,b). Then (u_0, u_1, u_2, f) constitutes a Tchebycheff system over $[a,b]$.*

Proof. First observe that if $\eta^2 f''$ is injective, then u_0, u_1, u_2 and f are linearly independent. Indeed, assume on the contrary that u_0, u_1, u_2 and f are not linearly independent. Then it holds that $f = a_0 u_0 + a_1 u_1 + a_2 u_2$ with $a_0, a_1, a_2 \in \mathbb{R}$ and $\sum_{i=0}^{2} a_i^2 > 0$, because u_0, u_1 and u_2 are linearly independent (cf. Lemma V.2.2). Hence, for all $x \in (a,b)$

$$\eta^2(x) f''(x) = \frac{2 f''(x)}{q_y''(x)} = 2a_2.$$

This contradicts the fact that $\eta^2 f''$ is injective.

Let $a \le x_0 < x_1 < x_2 < x_3 \le b$ and assume that

$$A := \begin{pmatrix} 1 & 1 & 1 & 1 \\ x_0 & x_1 & x_2 & x_3 \\ q_y(x_0) & q_y(x_1) & q_y(x_2) & q_y(x_3) \\ f(x_0) & f(x_1) & f(x_2) & f(x_3) \end{pmatrix}$$

is not invertible, i.e. $\det(A) = 0$. In particular, the system of linear equations

$$(w_0, w_1, w_2, w_3) A = (0,0,0,0)$$

has a non-trivial solution $(z_0, z_1, z_2, z_3) \in \mathbb{R}^4$. Now define

$$P(x) = z_0 + z_1 x + z_2 q_y(x) + z_3 f(x).$$

Then P has at least the four different roots x_0, x_1, x_2, x_3. Using Rolle's theorem, we conclude that P'' possesses at least two distinct roots in (a,b). Observe that

$$P''(x) = \frac{2 z_2}{\eta^2(x)} + z_3 f''(x).$$

If $z_3 \neq 0$, then any root s of P'' satisfies

$$\eta^2(s) f''(s) = -\frac{2 z_2}{z_3}. \tag{V.2.4}$$

Since $\eta^2 f''$ is injective, (V.2.4) has at most one solution. Hence, $z_3 = 0$. If $z_2 \neq 0$, then P'' has no root. Thus, it holds that $z_2 = 0$. Moreover, since P has at least the four different roots x_0, x_1, x_2, x_3, we conclude that $P(x) = 0$ for all $x \in [a, b]$. This contradicts the linear independence of u_0, u_1, u_2 and f. Therefore, $\det(A) \neq 0$ for fixed $a \leq x_0 < x_1 < x_2 < x_3 \leq b$. Finally, since the mapping $(x_0, x_1, x_2, x_3) \mapsto D\begin{pmatrix} 0 & 1 & 2 & 3 \\ x_0 & x_1 & x_2 & x_3 \end{pmatrix}$ is continuous, the extended system (u_0, u_1, u_2, f) constitutes a Tchebycheff system over $[a, b]$. $\square$

Remark V.2.4. Let $\eta\colon [a, b] \to (0, \infty)$ be measurable such that η^2 is continuous, then it holds that $\frac{1}{\eta^2} \in L^1_{loc}\big((a, b)\big)$ and the SDE

$$dY_t = \eta(Y_t) dW_t, \quad Y_0 = y \in (a, b), \tag{V.2.5}$$

driven by a Brownian motion W, admits a weak solution (Y, W) that is unique in law. In particular, it holds that $q_y(x) = \int_y^x \int_y^w \frac{2}{\eta^2(u)} du\, dw$ for $x \in [a, b]$, cf. Remark III.1.2.

In the following we collect functions η and f such that (u_0, u_1, u_2, f) forms a Tchebycheff system.

Example V.2.5.

a) Let $-\infty < a < b < \infty$ and $\eta(x) = 1$ for all $x \in [a, b]$. For $y \in (a, b)$ we have $q_y(x) = (x - y)^2$, $x \in [a, b]$, and the solution Y to (V.2.5) is a Brownian motion starting in $y \in (a, b)$ under $\mathbb{P}^y$ which is absorbed at the boundary points a and b.

- Let $0 < a < b < \infty$ and $f\colon [a, b] \to \mathbb{R}$, $f(x) = x^p$, $p \in \mathbb{R}\backslash\{0, 1, 2\}$. Then $f''(x) = p(p-1)x^{p-2}$ is injective.
- For $f\colon [a, b] \to \mathbb{R}$, $f(x) = x^{2n+1}$ with $n \in \mathbb{N}$, $n \geq 1$, we have that $f''(x) = 2n(2n+1)x^{2n-1}$ is injective.
- Let $f\colon [a, b] \to \mathbb{R}$, $f(x) = e^x$. Observe that $f''(x) = e^x$ is injective.

b) Let $0 < a < b < \infty$ and $\eta(x) = x$, $x \in [a, b]$. Hence, for $y \in (a, b)$ we conclude that $q_y(x) = 2\left(\frac{x}{y} - 1\right) - 2\log\left(\frac{x}{y}\right)$, $x \in [a, b]$. The diffusion Y solving (V.2.5) is a geometric Brownian motion starting in y under $\mathbb{P}^y$, $y \in (a, b)$, and Y is absorbed at a and b.

- Let $f\colon [a, b] \to \mathbb{R}$, $f(x) = x^p$, $p \in \mathbb{R}\backslash\{0, 1\}$. Then $\eta^2(x)f''(x) = p(p-1)x^p$ is injective.
- For $f\colon [a, b] \to \mathbb{R}$, $f(x) = e^x$, note that $\eta^2(x)f''(x) = x^2 e^x$ is injective.

c) Let $0 < a < b < \infty$ and $\eta(x) = \sqrt{x}$, $x \in [a, b]$. For $y \in (a, b)$ it holds that $q_y(x) = 2x\log\left(\frac{x}{y}\right) - 2(x - y)$, $x \in [a, b]$.

- Let $f\colon [a, b] \to \mathbb{R}$, $f(x) = x^p$, $p \in \mathbb{R}\backslash\{0, 1\}$. We have that $\eta^2(x)f''(x) = p(p-1)x^{p-1}$ is injective.
- For $f\colon [a, b] \to \mathbb{R}$, $f(x) = e^x$, the function $\eta^2(x)f''(x) = xe^x$ is injective.
- Let $f(x) = \log(x)$, $x \in [a, b]$. Observe that $\eta^2(x)f''(x) = -\frac{1}{x}$ is injective on $[a, b]$.

d) Let $0 < a < b < 1$ and $c > 0$. For $\eta(x) = cx(1 - x)$, $x \in [a, b]$, we deduce that $q_y(x) = \frac{2}{c^2}\left((2x - 1)\log\left(\frac{x(1-y)}{(1-x)y}\right) + (x - y)\frac{1-2y}{y(1-y)}\right)$.

- For $f(x) = -\log(1 - x)$ it holds that $\eta^2(x)f''(x) = c^2x^2$ is injective on $x \in [a, b]$.
- Let $f(x) = x\log\left(\frac{x}{1-x}\right)$, $x \in [a, b]$. Then $\eta^2(x)f''(x) = c^2 x$ is injective on $[a, b]$.

If the payoff function f and the function q_y are smooth enough, then the condition of Lemma V.2.3 is also necessary.

Lemma V.2.6. *Let $-\infty < a < b < \infty$ and let $\eta\colon [a,b] \to \mathbb{R}\backslash\{0\}$ be measurable. For $y \in (a,b)$ define $q_y(x) = \int_y^x \int_y^w \frac{2}{\eta^2(u)} du\, dw$, $x \in [a,b]$, and $q_y(x) = \infty$, $x \notin [a,b]$. Let $f\colon [a,b] \to \mathbb{R}$ and assume that $q_y, f \in \mathcal{C}^3\big((a,b)\big)$. If (u_0, u_1, u_2, f) constitutes a Tchebycheff system, then $\eta^2 f''$ is injective.*

Proof. We assume without loss of generality that the determinant in (V.1.1) is strictly positive. Otherwise we consider $-f$ instead of f. Since (u_0, u_1, u_2, f) is a Tchebycheff system, it holds for all $x \in (a,b)$ and all $h \in \big(0, \frac{b-x}{3}\big)$ that

$$
\begin{aligned}
0 &< \frac{1}{h^6} D \begin{pmatrix} 0 & 1 & 2 & 3 \\ x & x+h & x+2h & x+3h \end{pmatrix} \\
&= \left(\frac{f(x) - 3f(x+h) + 3f(x+2h) - f(x+3h)}{h^3} \right) \left(\frac{-q_y(x+h) + 2q_y(x+2h) - q_y(x+3h)}{h^2} \right) \\
&\quad + \left(\frac{f(x+h) - 2f(x+2h) + f(x+3h)}{h^2} \right) \left(\frac{q_y(x) - 3q_y(x+h) + 3q_y(x+2h) - q_y(x+3h)}{h^3} \right).
\end{aligned}
$$

Since $f, q_y \in \mathcal{C}^3\big((a,b)\big)$ L'Hôspital's rule implies

$$
\begin{aligned}
&\lim_{h \downarrow 0} \frac{f(x+h) - 2f(x+2h) + f(x+3h)}{h^2} = f''(x), \\
&\lim_{h \downarrow 0} \frac{f(x) - 3f(x+h) + 3f(x+2h) - f(x+3h)}{h^3} = -f'''(x), \\
&\lim_{h \downarrow 0} \frac{-q_y(x+h) + 2q_y(x+2h) - q_y(x+3h)}{h^2} = -q_y''(x), \\
&\lim_{h \downarrow 0} \frac{q_y(x) - 3q_y(x+h) + 3q_y(x+2h) - q_y(x+3h)}{h^3} = -q_y'''(x).
\end{aligned}
$$

Therefore, it holds for all $x \in (a,b)$ that

$$
0 \le f'''(x) q_y''(x) - f''(x) q_y'''(x). \tag{V.2.6}
$$

Since $\eta(x) > 0$, $x \in [a,b]$, by assumption, we conclude that $q_y''(x) = \frac{2}{\eta^2(x)} \in (0,\infty)$ for all $x \in (a,b)$. Hence, the necessary condition (V.2.6) combined with

$$
f'''(x) q_y''(x) - f''(x) q_y'''(x) = \big(q_y''(x)\big)^2 \left(\frac{f''}{q_y''} \right)'(x)
$$

yields

$$
\left(\frac{f''}{q_y''} \right)'(x) = \frac{1}{2} \left(\eta^2(x) f'' \right)'(x) \ge 0, \quad x \in (a,b).
$$

Assume that there exists an interval $(\underline{x}, \overline{x}) \subseteq (a,b)$ such that $\left(\frac{f''}{q_y''} \right)' \equiv 0$ on $(\underline{x}, \overline{x})$. In particular, there exists $c \in \mathbb{R}$ such that for all $x \in (\underline{x}, \overline{x})$

$$
\frac{f''(x)}{q_y''(x)} = c.
$$

Hence, the function f is given by

$$
f(x) = c q_y(x) + c_1 x + c_2
$$

on $(\underline{x}, \overline{x})$, where $c_1 = f'(\underline{x}) - cq_y'(\underline{x})$ and $c_2 = f(\underline{x}) - cq_y(\underline{x}) - \underline{x}[f'(\underline{x}) - cq_y'(\underline{x})]$. Choose $\underline{x} < x_0 < x_1 < x_2 < x_3 < \overline{x}$. Then

$$D\begin{pmatrix} 0 & 1 & 2 & 3 \\ x_0 & x_1 & x_2 & x_3 \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & 1 & 1 & 1 \\ x_0 & x_1 & x_2 & x_3 \\ q_y(x_0) & q_y(x_1) & q_y(x_2) & q_y(x_3) \\ cq_y(x_0) + c_1x_0 + c_2 & cq_y(x_1) + c_1x_1 + c_2 & cq_y(x_2) + c_1x_2 + c_2 & cq_y(x_3) + c_1x_3 + c_2 \end{pmatrix} = 0,$$

which contradicts the fact that (u_0, u_1, u_2, f) constitutes a Tchebycheff system. Therefore, there exists no interval $(\underline{x}, \overline{x}) \subseteq (a, b)$ with $\left(\frac{f''}{q_y''}\right)' \equiv 0$ on $(\underline{x}, \overline{x})$. To summarize, $\eta^2 f'' = \frac{2f''}{q_y''}$ is injective on $[a, b]$. $\square$

Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function such that (u_0, u_1, u_2, f) also forms a Tchebycheff system. Let $y \in (a, b)$ and $T \in (0, \infty)$ such that $\mathbf{c}^0 := \big(1, y, T\big) \in \operatorname{int}(\mathscr{M}_3)$. Then the value of the measure optimization problem (V.2.3) is given by

$$\widetilde{U}(T, y) = \sup_{\varsigma \in R((1,y,T))} \int_a^b f(x) d\varsigma(x).$$

We first derive the upper and lower principal representation of $\mathbf{c}^0$ and then apply Theorem V.1.25. Since the principal representations of $\mathbf{c}^0$ have index $\frac{3}{2}$, they can be represented by

$$\begin{aligned} \overline{\varsigma}(x) &= \gamma_1 \mathbb{1}_{[s^*,b]}(x) + \gamma_2 \mathbb{1}_{\{b\}}(x), \\ \underline{\varsigma}(x) &= \lambda_1 \mathbb{1}_{[a,b]}(x) + \lambda_2 \mathbb{1}_{[t^*,b]}(x), \end{aligned}$$

where $\gamma_j, \lambda_j > 0$, $j \in \{1, 2\}$, and $a < s^* < t^* < b$. Using $c_i^0 = \int_a^b u_i(x) d\overline{\varsigma}(x)$, $0 \leq i \leq 2$, we obtain the following system of equations

$$\begin{pmatrix} 1 & 1 \\ s^* & b \\ q_y(s^*) & q_y(b) \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} 1 \\ y \\ T \end{pmatrix}. \tag{V.2.7}$$

Therefore, we conclude that

$$\gamma_1 \in \left(\frac{b-y}{b-a}, 1\right), \qquad \gamma_2 = 1 - \gamma_1, \qquad s^* = b - \frac{b-y}{\gamma_1}.$$

In order to determine γ_1, we consider the last equation in (V.2.7), namely,

$$\gamma_1 q_y\left(b - \frac{b-y}{\gamma_1}\right) + (1 - \gamma_1) q_y(b) = T$$

and define the function $k\colon \left(\frac{b-y}{b-a}, 1\right) \to \mathbb{R}$, $k(\gamma) = \gamma q_y\left(b - \frac{b-y}{\gamma}\right) + (1-\gamma) q_y(b)$. Then

$$\begin{aligned} \lim_{\gamma \uparrow 1} k(\gamma) &= 0, \\ \lim_{\gamma \downarrow \frac{b-y}{b-a}} k(\gamma) &= \frac{b-y}{b-a} q_y(a) + \frac{y-a}{b-a} q_y(b). \end{aligned}$$

Moreover, k is strictly decreasing on $\left(\frac{b-y}{b-a}, 1\right)$. Indeed, let $\gamma, \lambda \in \left(\frac{b-y}{b-a}, 1\right)$ with $\gamma > \lambda$. Then

$$\begin{aligned}
k(\gamma) - k(\lambda) &= (\lambda - \gamma)\left[q_y(b) + \frac{1}{2}(b-y)\, m(\{y\})\right] + \gamma \int_{b-\frac{b-y}{\gamma}}^{y} m\big((u,y)\big)du - \lambda \int_{b-\frac{b-y}{\lambda}}^{y} m\big((u,y)\big)du \\
&= (\lambda - \gamma)\left[q_y(b) + \frac{1}{2}(b-y)m(\{y\})\right] + (\gamma - \lambda)\int_{b-\frac{b-y}{\gamma}}^{y} m\big((u,y)\big)du - \lambda \int_{b-\frac{b-y}{\lambda}}^{b-\frac{b-y}{\gamma}} m\big((u,y)\big)du \\
&< (\lambda - \gamma)\left[q_y(b) + \frac{1}{2}(b-y)\, m(\{y\}) + (b-y)\, m\left(\left(b - \frac{b-y}{\gamma}, y\right)\right)\right] \\
&< 0.
\end{aligned}$$

As a consequence, for every $T \in \left(0, \frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b)\right)$ there exists a unique $\gamma^* = \gamma^*(T,y) \in \left(\frac{b-y}{b-a}, 1\right)$ such that $k(\gamma^*) = T$. Repeating the calculation for $\underline{\varsigma}$ and using Theorem V.1.25 and Theorem V.1.17 we obtain the following theorem.

Theorem V.2.7. *Let $y \in (a,b)$. Let $f\colon [a,b] \to \mathbb{R}$ be continuous such that (u_0, u_1, q_y, f) constitutes a Tchebycheff system over $[a,b]$.*

1. *It holds that $\mathbf{c}^0 = (1, y, T) \in int(\mathscr{M}_3)$ if and only if $T \in \left(0, \frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b)\right)$. In this case the upper principal representation $\overline{\varsigma}$ and the lower principal representation $\underline{\varsigma}$ are given by*

$$\begin{aligned}
\overline{\varsigma}(x) &= \gamma^* \mathbb{1}_{\left[b-\frac{b-y}{\gamma^*}, b\right]}(x) + (1-\gamma^*)\mathbb{1}_{\{b\}}(x), \\
\underline{\varsigma}(x) &= (1-\lambda^*)\mathbb{1}_{[a,b]}(x) + \lambda^* \mathbb{1}_{\left[a+\frac{y-a}{\lambda^*}, b\right]}(x),
\end{aligned}$$

where $\gamma^* = \gamma^*(T,y) \in \left(\frac{b-y}{b-a}, 1\right)$ is the unique solution of $\gamma q_y\left(b - \frac{b-y}{\gamma}\right) + (1-\gamma)q_y(b) = T$ and $\lambda^* = \lambda^*(T,y) \in \left(\frac{y-a}{b-a}, 1\right)$ is the unique solution of $(1-\lambda)q_y(a) + \lambda q_y\left(a + \frac{y-a}{\lambda}\right) = T$. Moreover, we have

$$\begin{aligned}
\widetilde{U}(T,y) &= \sup_{\varsigma \in R(\mathbf{c}^0)} \int_a^b f(x)d\varsigma(x) \\
&= \max\left\{\gamma^*(T,y) f\left(b - \frac{b-y}{\gamma^*(T,y)}\right) + \big(1-\gamma^*(T,y)\big)f(b),\right. \\
&\qquad\quad \left.\big(1-\lambda^*(T,y)\big)f(a) + \lambda^*(T,y) f\left(a + \frac{y-a}{\lambda^*(T,y)}\right)\right\}.
\end{aligned}$$

2. *If $T = \frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b)$, then $R(\mathbf{c}^0) = \{\varsigma^*\}$, where $\varsigma^*(x) = \frac{b-y}{b-a}\mathbb{1}_{[a,b]}(x) + \frac{y-a}{b-a}\mathbb{1}_{\{b\}}(x)$ and*

$$\widetilde{U}\left(\frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b), y\right) = \frac{b-y}{b-a}f(a) + \frac{y-a}{b-a}f(b).$$

Remark V.2.8. Let $y \in (a,b)$. If f is a continuous function on $[a,b]$ such that (u_0, u_1, q_y, f) is a Tchebycheff system and the determinant (V.1.3) is strictly positive, then for $T \in \left(0, \frac{b-y}{b-a}q_y(a) + \frac{y-a}{b-a}q_y(b)\right)$ we have

$$\widetilde{U}(T,y) = \gamma^*(T,y) f\left(b - \frac{b-y}{\gamma^*(T,y)}\right) + \big(1-\gamma^*(T,y)\big)f(b),$$

where $\gamma^*(T,y)$ is the unique solution of $\gamma q_y\left(b - \frac{b-y}{\gamma}\right) + (1-\gamma)q_y(b) = T$, see Theorem V.1.25. Conversely, if the determinant (V.1.3) is strictly negative, then the unique optimal cumulative distribution function is given by the lower principal representation $\underline{\varsigma}$.

Theorem V.2.7 implies that in the measure optimization problem (V.2.2) it is sufficient to consider only measures in $\mathcal{A}_2(T,y)$. More precisely, we can restrict to measures of the form

$$\mu^*(S,y) = \gamma^*(S,y)\delta_{b-\frac{b-y}{\gamma^*(S,y)}} + \big(1-\gamma^*(S,y)\big)\delta_b,$$
$$\mu_*(S,y) = \big(1-\lambda^*(S,y)\big)\delta_a + \lambda^*(S,y)\delta_{a+\frac{y-a}{\lambda^*(S,y)}}$$

with $S \in [0,T]$. The stopping times $\tau^*(S,y)$ and $\tau_*(S,y)$ embedding μ^* and μ_* in Y under $\mathbb{P}^y$, respectively, are the first exit times of the intervals $\Big(b - \frac{b-y}{\gamma^*(S,y)}, b\Big)$ and $\Big(a, a + \frac{y-a}{\lambda^*(S,y)}\Big)$, respectively, see the proof of Theorem III.1.5.

Remark V.2.9. For a general diffusion Y with compact state space, Theorem V.2.7 provides a class of payoff functions f such that the stopping time $\tau^*(S,y)$ is optimal for the stopping problem (V.2.1).

Corollary V.2.10. *Let $y \in (a,b)$ and $T \in [0,\infty)$. If (u_0,u_1,u_2,f) is a Tchebycheff system over $[a,b]$, then 2 points suffice in* (V.2.1), *i.e.*

$$V(T,y) = \sup_{\mu\in\mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

Moreover, for the optimal stopping problem (V.2.2) *it holds that*

$$V(T,y) = \sup_{\tau\in\mathcal{S}_2(T,y)} \mathbb{E}^y[f(Y_\tau)],$$

where $\mathcal{S}_2(T,y)$ denotes the set of stopping times $\tau_{a_1,b_1} = \inf\{t \in [0,\infty)\colon Y_t \notin (a_1,b_1)\}$, where $a \le a_1 \le b_1 \le b$ such that $\frac{b_1-y}{b_1-a_1}q_y(a_1) + \frac{y-a_1}{b_1-a_1}q_y(b_1) \le T$.

We close this section by focusing on processes Y which have non-compact state space J with interior (a,b), $-\infty \le a < b \le \infty$. We assume that Y satisfies the assumptions from Chapter III. Here we describe the main steps which allow to restrict to first exit times in (V.2.1). First observe that for every $\mu \in \mathcal{A}_3(T,y)$ there exists a compact set $J_\mu \subseteq J$ such that the support of μ is contained in J_μ. Since a reduction to weighted sums of three Dirac measures is possible by Theorem III.2.5, it holds that

$$\begin{aligned} V(T,y) &= \sup_{\mu\in\mathcal{A}_3(T,y)} \int_{\mathbb{R}} f(x)\mu(dx) = \sup_{z\in J}\sup_{k\in\mathbb{N}}\sup_{\mu\in\mathcal{A}_3^k(T,y;z)} \int_{\mathbb{R}} f(x)\mu(dx) \\ &= \sup_{z\in J}\sup_{k\in\mathbb{N}}\sup_{\mu\in\mathcal{A}^k(T,y;z)} \int_{\mathbb{R}} f(x)\mu(dx), \end{aligned} \tag{V.2.8}$$

where

$$\mathcal{A}^k(T,y;z) = \left\{\mu \in \mathcal{A}(T,y)\colon \mu\big([a_k,b_k]\big) = 1, \int_{\mathbb{R}} x\,\mu(dx) = z\right\},$$
$$\mathcal{A}_3^k(T,y;z) = \mathcal{A}^k(T,y;z) \cap \mathcal{A}_3(T,y)$$

and $(a_k)_{k\in\mathbb{N}}, (b_k)_{k\in\mathbb{N}}$ are sequences in J with $a_k < y < b_k$ for $k \in \mathbb{N}$, $a_k \searrow a$ and $b_k \nearrow b$ as $k \to \infty$. For all $k \in \mathbb{N}$ we set $a_k = a$ if $l \in J$ and $b_k = b$ if $r \in J$. Since in general the measures in $\mathcal{A}(T,y)$ are not centered around y if J is not bounded, we consider the supremum over all $z \in J$ in (V.2.8). Here we set $\sup\emptyset = -\infty$.

Assume that the functions $u_0(x) = 1$, $u_1(x) = x$, $u_2(x) = q_y(x)$ and f constitute a Tchebycheff system over $[a_k,b_k]$ for all $k \in \mathbb{N}$. Let $k \in \mathbb{N}$ and $z \in J$ such that $\mathcal{A}^k(T,y;z) \neq \emptyset$. Then Theorem V.1.25 implies that we can confine to weighted sums of at most two Dirac measures in the optimization problem

$$\sup_{\mu\in\mathcal{A}^k(T,y;z)} \int_{\mathbb{R}} f(x)\mu(dx). \tag{V.2.9}$$

The measure optimization (V.2.9) corresponds to an optimal stopping problem for the process $Y_t^k = Y_{t\wedge\tau_{a_k,b_k}}$, $t \in [0,\infty)$, over stopping times τ with $\mathbb{E}^y[\tau] \le T$ and $\mathbb{E}^y[Y_\tau] = z$.

To summarize, the following theorem holds.

Theorem V.2.11. *Let $(Y_t)_{t\in[0,\infty)}$ be a general diffusion with state space J, where the interior of J is given by (a,b), $-\infty \le a < b \le \infty$. Let $y \in (a,b)$ and let $f : J \to \mathbb{R}$ be continuous such that (u_0, u_1, q_y, f) forms a Tchebycheff system over $[a_k, b_k]$ for every $k \in \mathbb{N}$. Then*

$$V(T,y) = \sup_{\mu\in\mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

In addition, it is sufficient to focus on first exit times in (V.2.1)*, i.e.*

$$V(T,y) = \sup_{\tau\in\mathcal{S}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

Remark V.2.12. In Example II.4.6 we consider a Brownian motion on $\mathbb{R}$ and the payoff function $f(x) = |x|$. The optimal stopping time τ^* for $V(T,y)$ is the first exit time of the interval $\big(-\sqrt{T+y^2}, \sqrt{T+y^2}\big)$. Thus, the law μ^* of the Brownian motion stopped at τ^* under $\mathbb{P}^y$ is given by

$$\mu^* = \left(\frac{1}{2} - \frac{y}{2\sqrt{T+y^2}}\right)\delta_{-\sqrt{T+y^2}} + \left(\frac{1}{2} + \frac{y}{2\sqrt{T+y^2}}\right)\delta_{\sqrt{T+y^2}}.$$

The measure μ^* is optimal in (V.2.1) by Corollary III.2.5.

Observe that the extended system $(u_i)_{i=0}^3$ with $u_0(x) = 1$, $u_1(x) = x$, $u_2(x) = q_y(x) = (x-y)^2$, $u_3(x) = |x|$ does not form a Tchebycheff system over any interval $[a_1, b_1]$, $a_1, b_1 \in \mathbb{R}$, $a_1 < b_1$. Indeed, for $a_1 b_1 \ge 0$ the functions $(u_i)_{i=0}^3$ are not linearly independent and if $a_1 < 0 < b_1$, then for $a_1 \le x_0 < x_1 < x_2 < x_3 \le b_1$ we obtain

$$D\begin{pmatrix} 0 & 1 & 2 & 3 \\ x_0 & x_1 & x_2 & x_3 \end{pmatrix} = \begin{cases} 2x_0(x_2-x_1)(x_3-x_1)(x_3-x_2) < 0, & \text{if } x_0 < 0 \le x_1, \\ 2x_3(x_1-x_0)(x_2-x_0)(x_2-x_1) > 0, & \text{if } x_2 \le 0 < x_3. \end{cases}$$

Thus, we cannot apply Theorem V.2.7 and the succeeding discussion to conclude that we can restrict to first exit times in (V.2.1).

V.2.2 Brownian Motion on $[a,b]$

Let $-\infty < a < b < \infty$. We consider the case of a Brownian motion starting in y under $\mathbb{P}^y$, $y \in (a,b)$, which is absorbed at a and b. Recall that $q_y(x) = (x-y)^2$, $x \in [a,b]$. Let $f\colon [a,b] \to \mathbb{R}$ be continuous such that (u_0, u_1, q_y, f) is a Tchebycheff system. Let $\mathbf{c}^0 = (1, y, T)$ with $T \in \big(0, (b-y)(y-a)\big]$. Then the optimization problem (V.2.3) simplifies to

$$\begin{aligned}\widetilde{U}(T,y) = \sup_{\varsigma\in R(\mathbf{c}^0)} \int_a^b f(x)d\varsigma(x) = \max\bigg\{ &\frac{(b-y)^2}{T+(b-y)^2} f\left(y - \frac{T}{b-y}\right) + \frac{T}{T+(b-y)^2} f(b)\,, \\ &\frac{(y-a)^2}{T+(y-a)^2} f\left(y + \frac{T}{y-a}\right) + \frac{T}{T+(y-a)^2} f(a)\bigg\}.\end{aligned}$$

If $T = (b-y)(y-a)$, then $R(\mathbf{c}^0) = \{\varsigma^*\}$, where $\varsigma^*(x) = \frac{b-y}{b-a}\mathbb{1}_{[a,b]}(x) + \frac{y-a}{b-a}\mathbb{1}_{\{b\}}(x)$. In particular, by Corollary (V.2.10) it holds that

$$V(T,y) = \sup_{\mu\in\mathcal{A}_2(T,y)} \int_{\mathbb{R}} f(x)\mu(dx).$$

Now we specify some functions f for which we can compute the determinant explicitly, such that the extended system (u_0, u_1, u_2, f) is also a Tchebycheff system. For further functions we refer to Example V.2.5 a).

Example V.2.13.
a) Let $f\colon [a,b] \to \mathbb{R}$, $0 \le a < b$, $f(x) = \sqrt{x}$. Then we obtain for all $a \le x_0 < x_1 < x_2 < x_3 \le b$

$$D\begin{pmatrix} 0 & 1 & 2 & 3 \\ x_0 & x_1 & x_2 & x_3 \end{pmatrix} = (\sqrt{x_0} + \sqrt{x_1} + \sqrt{x_2} + \sqrt{x_3}) \prod_{0 \le i < j \le 3} \left(\sqrt{x_j} - \sqrt{x_i}\right) > 0.$$

b) For $f\colon [a,b] \to \mathbb{R}$, $0 \le a < b$, $f(x) = x^{\frac{3}{2}}$ and all $a \le x_0 < x_1 < x_2 < x_3 \le b$ it holds that

$$D\begin{pmatrix} 0 & 1 & 2 & 3 \\ x_0 & x_1 & x_2 & x_3 \end{pmatrix} = -\left(\sqrt{x_0x_1x_2} + \sqrt{x_0x_1x_3} + \sqrt{x_0x_2x_3} + \sqrt{x_1x_2x_3}\right) \prod_{0 \le i < j \le 3} \left(\sqrt{x_j} - \sqrt{x_i}\right) < 0.$$

c) Let $f\colon [a,b] \to \mathbb{R}$, $f(x) = \frac{1}{x}$, where $ab > 0$ and $a \le x_0 < x_1 < x_2 < x_3 \le b$. Then,

$$D\begin{pmatrix} 0 & 1 & 2 & 3 \\ x_0 & x_1 & x_2 & x_3 \end{pmatrix} = -\frac{(x_1 - x_0)(x_2 - x_0)(x_2 - x_1)(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)}{x_0x_1x_2x_3} < 0.$$

In the following example we derive the value of the optimal stopping problem (V.2.1) for $f(x) = x^3$.

Example V.2.14. Let Y be a Brownian motion on $[a,b]$, $a, b \in \mathbb{R}$, $a < b$ and $f(x) = x^3$. For $a < y < b$ and $T \in \big(0, (b-y)(y-a)\big]$ we conclude from Theorem V.2.7 that

$$\begin{aligned} V(T,y) &= \sup_{S \in [0,T]} \max\left\{ bS + 2Sy + y^3 - \frac{S^2}{b-y}, aS + 2Sy + y^3 + \frac{S^2}{y-a} \right\} \\ &= \sup_{S \in [0,T]} \left(bS + 2Sy + y^3 - \frac{S^2}{b-y} \right). \end{aligned}$$

Remark V.2.1b) implies that we can compute $V(T,y)$ for all $T \in [0,\infty)$. For this purpose we distinguish three different cases: If $b \le 0$, then the payoff function is concave on $[a,b]$ and stopping directly is optimal. Hence, $V(T,y) = y^3$ for all $T \in [0,\infty)$.

If $b > 0$ and $b + 2a < 0$, it holds that

$$V(T,y) = \begin{cases} y^3, & y \le -\frac{b}{2}, \\ bT + 2yT + y^3 - \frac{T^2}{b-y}, & y > -\frac{b}{2},\, T < \frac{1}{2}(b-y)(b+2y), \\ \frac{1}{4}b^2(b+3y), & y > -\frac{b}{2},\, T \ge \frac{1}{2}(b-y)(b+2y). \end{cases}$$

Finally, if $b + 2a \ge 0$, then the value function V is given by

$$V(T,y) = \begin{cases} bT + 2yT + y^3 - \frac{T^2}{b-y}, & T < (b-y)(y-a), \\ (a^2 + ab + b^2)y - ab(a+b), & T \ge (b-y)(y-a). \end{cases}$$

VI. Conclusion

We summarize the results of this thesis and indicate possible extensions.

In Chapter II we analyze optimal stopping problems with expectation cost constraints, where the underlying process is a solution to a time-homogeneous SDE driven by a Brownian motion. By extending the state space with the process of the conditional expected constraint, we transform the constrained stopping problem into an unconstrained control problem. Thus, we can formulate a DPP and derive a DPE for the value function of the control problem. The main result is a verification theorem.

We focus on time-homogeneous SDEs, constraint functions h and payoff functions f. One could investigate optimal stopping problems for time-inhomogeneous processes and functions as well as multiple constraints of the form

$$\mathbb{E}\left[\int_0^\tau h^i(s, X_s)ds + g^i(\tau, X_\tau)\right] \leq T_i, \qquad 1 \leq i \leq n, \tag{VI.1}$$

where $g^i, h^i \colon [0,\infty) \times \mathbb{R}^n \to (0,\infty]$ are measurable and $T_i \in [0,\infty)$. Besides expectation cost constraints, the constraints (VI.1) also cover distribution constraints on the stopping time τ and on X_τ. To establish a one-to-one correspondence similar to Proposition II.2.4 one has to carefully deal with the time-dependence of g^i and h^i.

Furthermore, one could try to solve the PDE (II.3.3) numerically. Related is the question whether the value function U is a viscosity solution to (II.3.3) and whether there is a comparison principle for viscosity solutions. If the underlying process is a one-dimensional Brownian motion and $h(x) = 1$, $x \in \mathbb{R}$, then Miller [43] shows that U is the unique viscosity solution to (II.3.3). He also establishes a connection to Monge-Ampère equations which are non-linear second order PDEs of the form

$$\det(D^2u) = k(x, u, Du), \quad u \text{ convex},$$

where $k(x, r, p) \geq 0$, see Example 1.11 in [22]. In our setting, if $d = n = 1$ and if U is strictly concave in T, then (II.3.3) can be rewritten as

$$-\det\begin{pmatrix} U_{TT}(T,x) & U_{Tx}(T,x) \\ U_{Tx}(T,x) & U_{xx}(T,x) - \dfrac{h(x)U_T(T,x) - b(x)U_x(T,x)}{\sigma^2(x)} \end{pmatrix} = 0.$$

The PDE can be interpreted as a Monge-Ampère type equation. Thus, one could try to apply numerical schemes for Monge-Ampère equations, consult e.g. [46].

Chapter III shows that for optimally stopping a one-dimensional regular continuous strong Markov process Y with stopping times τ satisfying the expectation constraint $\mathbb{E}[\tau] \leq T$, we can confine to stopping times τ such that the law of Y_τ is a weighted sum of at most 3 Dirac measures. Here we exploit the fact that Skorokhod embedding techniques allow to characterize all measures that can be embedded in Y with a stopping time satisfying the expectation constraint.

It would be interesting to consider further constraints, e.g. expectation cost constraints or probability constraints, as well as n-dimensional processes and to investigate whether we can reduce the set of stopping times. Here, the main task is to transform the optimal stopping problem into a measure optimization problem.

In Chapter IV we study a sequential testing problem and illustrate the results obtained in Chapter II and III. By continuously observing a Brownian motion having either drift 0 or κ, $\kappa \neq 0$, an economic agent wants to detect the value of the drift. An exogenously given threshold $\alpha \in \left(0, \frac{1}{2}\right)$ allows the agent to accept a drift of 0 and κ at time $t \in [0, \infty)$ only if the a posteriori probability process Y at time t is less or equal to α or exceeds $1 - \alpha$, respectively. We show that if we impose an expectation constraint on the time until the agent has to terminate observing the Brownian motion, then it can be optimal to end with no significant result. In this case, the optimal stopping time is a composition of two exit times.

One could examine how the value function and the optimal stopping time change if we allow for $\alpha \in (0, 1)$ and modify the acceptance rule as follows: Accept the drift 0 and κ at time t only if $Y_t \leq \alpha$ and $Y_t \geq 1 - \gamma$, $\gamma \in (0, 1 - \alpha)$, respectively.

Finally, Chapter V focuses on reducing the stopping times in the optimal stopping problem presented in Chapter III to first exit times. We apply results from the theory of Tchebycheff systems to the measure optimization problem from Chapter III in order to characterize the maximizer μ^* for $\int_{\mathbb{R}} f(x)\mu(dx)$ over all probability measures $\mu \in \mathcal{A}(T, y)$ provided that $\left(1, x, q_y(x), f(x)\right)$ forms a Tchebycheff system. The optimal measure μ^* is a weighted sum of 2 Dirac measures and the stopping time embedding μ^* into Y is thus a first exit time. This allows to confine to first exit times in the optimal stopping problem. We provide an example where the optimal stopping time is a first exit time but (u_0, u_1, u_2, f) is not a Tchebycheff system.

It would be worth deriving weaker conditions on the payoff function f that allow a reduction to first exit times in the optimal stopping problem.

A. Appendix

In this appendix we prove auxiliary statements for the results in Chapter II – V.

A.1 Appendix: Optimal Stopping with Expectation Constraints

The following lemma provides sufficient conditions for the value function U to be a solution of the DPE (II.3.3) on the whole set $(0,\infty)\times\mathbb{R}^n$ (cf. Remark II.3.5a)).

Lemma A.1.1. *Let h be continuous. Assume that $U\in\mathcal{C}^2\big((0,\infty)\times\mathbb{R}^n\big)$ and that U satisfies the DPP* (II.3.2)*. If, in addition, $\left|\sigma^\top\cdot\nabla_x U_T\right|^2/U_{TT}$ is continuous on $(0,\infty)\times\mathbb{R}^n$ and there exists an optimal control for every $(T,x)\in(0,\infty)\times\mathbb{R}^n$, then U is a solution to* (II.3.3) *on $(0,\infty)\times\mathbb{R}^n$.*

Proof. Let $(T,x)\in(0,\infty)\times\mathbb{R}^n$ and α^* be an optimal control for $U(T,x)$ with localizing sequence $(\tau_n)_{n\in\mathbb{N}}$. Set $X=X^x$, $H=H^x$ and $M^*=M^{\alpha^*,T}$. Define

$$\theta=\inf\left\{t\in[0,\infty)\colon M_t^*-H_t\notin\left[\frac{T}{2},2T\right]\text{ or }|X_t-x|\geq 1\right\}.$$

Since α^* is optimal, similar arguments as in the derivation of (II.3.4) entail that for $t\in(0,\infty)$ we have

$$\mathbb{E}\left[\int_0^{t\wedge\theta\wedge\tau_n}\left(-h(X_s)U_T+\mathcal{L}U+\frac{|\alpha_s^*|^2}{2}U_{TT}+(\nabla_x U_T)^\top\cdot\sigma(X_s)\cdot(\alpha_s^*)^\top\right)(M_s^*-H_s,X_s)ds\right]=0. \tag{A.1.1}$$

Observe that the integrand is non-positive, because U is a supersolution to (II.3.12) on $(0,\infty)\times\mathbb{R}^n$ by Proposition II.3.4 and Remark II.3.7. Hence, for $\mathbb{P}$ - a.a. $\omega\in\Omega$ and Lebesgue - a.a. $s\in[0,t\wedge\theta(\omega)\wedge\tau_n(\omega)]$ it follows that

$$\left(-h\big(X_s(\omega)\big)U_T+\mathcal{L}U+\frac{|\alpha_s^*(\omega)|^2}{2}U_{TT}+(\nabla_x U_T)^\top\cdot\sigma\big(X_s(\omega)\big)\cdot\big(\alpha_s^*(\omega)\big)^\top\right)\big(M_s^*(\omega)-H_s(\omega),X_s(\omega)\big)=0$$

and thus, (II.3.12) implies

$$\alpha_s^*(\omega)\in\operatorname*{arg\,max}_{a\in\mathbb{R}^d}\left\{\left(\frac{|a|^2}{2}U_{TT}+(\nabla_x U_T)^\top\cdot\sigma\big(X_s(\omega)\big)\cdot a\right)\big(M_s^*(\omega)-H_s(\omega),X_s(\omega)\big)\right\}.$$

In particular, if $U_{TT}\big(M_s^*(\omega)-H_s(\omega),X_s(\omega)\big)<0$, then

$$\alpha_s^*(\omega)=-\frac{(\nabla_x U_T)^\top\cdot\sigma\big(X_s^x(\omega)\big)}{U_{TT}}\big(M_s^*(\omega)-H_s(\omega),X_s(\omega)\big).$$

If $U_{TT}\big(M_s^*(\omega)-H_s(\omega),X_s(\omega)\big)=0$, then also $\nabla_x U_T\big(M_s^*(\omega)-H_s(\omega),X_s(\omega)\big)=0$ by Proposition II.3.4. Therefore, (A.1.1) can be rewritten as

$$\mathbb{E}\left[\int_0^{t\wedge\theta\wedge\tau_n}\left(-h(X_s)U_T+\mathcal{L}U-\frac{\left|\sigma^\top(X_s)\cdot\nabla_x U_T\right|^2}{2U_{TT}}\right)(M_s^*-H_s,X_s)ds\right]=0.$$

Recall that $\left|\sigma^\top(x)\cdot\nabla_x U_T(T,x)\right|^2/U_{TT}(T,x)=0$ if both the numerator and the denominator equal 0. Due to the continuity of h and $\left|\sigma^\top\cdot\nabla_x U_T\right|^2/U_{TT}$ it follows as in the proof of Proposition II.3.4 that

$$h(x)U_T(T,x)-\mathcal{L}U(T,x)+\frac{\left|\sigma^\top(x)\cdot\nabla_x U_T(T,x)\right|^2}{2U_{TT}(T,x)}=0.$$

□

In the next lemma we perform a verification for Example II.4.7 although the growth condition (II.4.1) is not satisfied.

Lemma A.1.2. *Let $f(y)=-y^4+y^2$, $h(y)=1$, $y\in\mathbb{R}$, and $X_t^x=x+W_t$, $t\in[0,\infty)$, be a one-dimensional Brownian motion starting in $x\in\mathbb{R}$.*

1. *Let $\tau\in\mathcal{T}(T)$, then we have $\mathbb{E}[f(X_\tau^x)]>-\infty$ if and only if $\mathbb{E}[\tau^2]<\infty$.*
2. *The value of the optimal stopping problem* (II.1.3) *is attained in the set of all square integrable $\tau\in\mathcal{T}(T)$. More precisely,*

$$U(T,x)=\sup_{\tau\in\mathcal{T}(T)\,:\,\mathbb{E}[\tau^2]<\infty}\mathbb{E}[f(X_\tau^x)].$$

3. *Let $u(T,x)=-x^4+x^2+T-T^2-2x^2T$, then $u(T,x)\geq U(T,x)$ for all $(T,x)\in(0,\infty)\times\mathbb{R}^n$.*

Proof. 1. First observe that for $\tau\in\mathcal{T}(T)$ we have $\mathbb{E}[f(X_\tau^x)]=\mathbb{E}\big[-|X_\tau^x|^4+|X_\tau^x|^2\big]>-\infty$ if and only if $\mathbb{E}\big[|X_\tau^x|^4\big]<\infty$ by (II.4.8). Hence, we prove that $\mathbb{E}\big[|X_\tau^x|^4\big]<\infty$ if and only if $\mathbb{E}\big[\tau^2\big]<\infty$.

Let $\tau\in\mathcal{T}(T)$ such that $\mathbb{E}\big[|X_\tau^x|^4\big]<\infty$. Doob's L^4-inequality implies

$$\mathbb{E}\left[\sup_{0\leq s\leq\tau}|X_s^x|^4\right]\leq\left(\frac{4}{3}\right)^4\mathbb{E}\left[|X_\tau^x|^4\right]<\infty.$$

Observe that we cannot directly apply the Burkholder-Davis-Gundy inequality to the left-hand side in order to show that $\mathbb{E}[\tau^2]<\infty$, because in general $X_0^x=x\neq 0$. In the following we slightly modify the proof of the Burkholder-Davis-Gundy inequality as presented in [56] to show that there exists $C\in(0,\infty)$ such that

$$\mathbb{E}\big[\tau^2\big]\leq C\,\mathbb{E}\left[\sup_{0\leq s\leq\tau}|X_s^x|^4\right]. \tag{A.1.2}$$

Applying Itô's formula and Itô's isometry imply that for $n\in\mathbb{N}$

$$\begin{aligned}
\mathbb{E}\big[(\tau\wedge n)^2\big]&=\mathbb{E}\left[\left(|X_{\tau\wedge n}^x|^2-|x|^2-\int_0^{\tau\wedge n}2X_s^x\,dW_s\right)^2\right]\\
&\leq 3\mathbb{E}\left[|X_{\tau\wedge n}^x|^4+|x|^4+\left(\int_0^{\tau\wedge n}2X_s^x dW_s\right)^2\right]\\
&\leq 6\mathbb{E}\left[\sup_{0\leq s\leq\tau\wedge n}|X_s^x|^4\right]+12\mathbb{E}\left[\int_0^{\tau\wedge n}|X_s^x|^2 ds\right]\\
&\leq 6\mathbb{E}\left[\sup_{0\leq s\leq\tau}|X_s^x|^4\right]+12\mathbb{E}\left[(\tau\wedge n)^2\right]^{1/2}\mathbb{E}\left[\sup_{0\leq s\leq\tau}|X_s^x|^4\right]^{1/2}.
\end{aligned}$$

Therefore, we conclude that

$$\mathbb{E}\big[(\tau\wedge n)^2\big]\leq\big(6+\sqrt{42}\big)^2\mathbb{E}\left[\sup_{0\leq s\leq\tau}|X_s^x|^4\right].$$

The monotone convergence theorem now implies (A.1.2) and hence $\mathbb{E}[\tau^2] < \infty$.

For the reverse statement let $\tau \in \mathcal{T}(T)$ be square integrable. There exist $C, \hat{C} \in (0, \infty)$ such that

$$\mathbb{E}\left[\left|X_\tau^x\right|^4\right] \le \mathbb{E}\left[\sup_{0\le s\le\tau} \left|X_s^x\right|^4\right] \le C\left(|x|^4 + \mathbb{E}\left[\sup_{0\le s\le\tau}\left|W_s\right|^4\right]\right) \le \hat{C}\left(|x|^4 + \mathbb{E}\left[\tau^2\right]\right) \tag{A.1.3}$$

by the Burkholder-Davis-Gundy inequality. Hence, $\mathbb{E}\left[|X_\tau^x|^4\right]$ is integrable.

2. By the first part we have that $\mathbb{E}[f(X_\tau^x)] = -\infty$ for all $\tau \in \mathcal{T}(T)$ with $\mathbb{E}[\tau^2] = \infty$. Therefore,

$$U(T,x) = \sup_{\tau\in\mathcal{T}(T)} \mathbb{E}[f(X_\tau^x)] = \sup_{\tau\in\mathcal{T}(T)\,:\,\mathbb{E}[\tau^2]<\infty} \mathbb{E}[f(X_\tau^x)].$$

3. By the second part it is sufficient to show that $u(T,x) \ge \mathbb{E}[f(X_\tau^x)]$ for all square integrable $\tau \in \mathcal{T}(T)$. Let $\tau \in \mathcal{T}(T)$ be square integrable with corresponding control α. Observe that $\alpha \in L^2(W)$ by the martingale representation theorem and define

$$\theta_n = \inf\left\{t \in [0,\infty)\colon M_t^{\alpha,T} - t \notin \left[\frac{T}{n}, Tn\right] \text{ or } |X_t^x| \ge n\right\} \wedge n.$$

Then the same arguments as in the proof of the first part of Theorem II.4.1 entail that

$$u(T,x) \ge \mathbb{E}\left[u\big(M_{\theta_n}^{\alpha,T} - \theta_n, X_{\theta_n}^x\big)\right]$$

and

$$u\big(M_{\theta_n}^{\alpha,T} - \theta_n, X_{\theta_n}^x\big) \xrightarrow[n\to\infty]{} u\big(0, X_{\tau^{\alpha,T}}^x\big) = f\big(X_{\tau^{\alpha,T}}^x\big),$$

where $\tau^{\alpha,T} = \inf\left\{t \in [0,\infty)\colon M_t^{\alpha,T} \le H_t\right\}$. The function u satisfies the following growth condition

$$\left|u\big(M_{\theta_n}^{\alpha,T} - \theta_n, X_{\theta_n}^x\big)\right| \le C\left(1 + \left|M_{\theta_n}^{\alpha,T}\right|^2 + \left|X_{\theta_n}^x\right|^4\right).$$

Note that (A.1.3) implies that $\left\{|X_\vartheta^x|^4\colon \vartheta\ (\mathcal{F}_t)\text{-stopping time, } \vartheta \le \tau \text{ a.s.}\right\}$ is uniformly integrable. Moreover, $\left\{|M_\vartheta|^2\colon \vartheta\ (\mathcal{F}_t)\text{-stopping time, } \vartheta \le \tau \text{ a.s.}\right\}$ is uniformly integrable for all square integrable $\tau \in \mathcal{T}(T)$ by Doob's L^2-inequality. Thus, we deduce that

$$\lim_{n\to\infty} \mathbb{E}\left[u\big(M_{\theta_n}^{\alpha,T} - \theta_n, X_{\theta_n}^x\big)\right] = \mathbb{E}\left[f\big(X_{\tau^{\alpha,T}}^x\big)\right]$$

and hence $u(T,x) \ge U(T,x)$. □

A.2 Appendix: 3 Points Suffice

We prove in this section several statements which are used in Chapter III. In Lemma A.2.1 we obtain the distribution of the process Y stopped at the first exit time of an interval after a prescribed stopping time.

Lemma A.2.1. *Let ϑ be an integrable $(\mathcal{F}_t)$-stopping time such that $(Y_{t\wedge\vartheta})_{t\in[0,\infty)}$ is bounded. Denote by ν the distribution of Y_ϑ under $\mathbb{P}^y$, i.e. $\mathbb{P}^y[Y_\vartheta \in A] = \nu(A)$ for all $A \in \mathcal{B}(\mathbb{R})$. For $a, b \in J$, $a < b$, let*

$$\rho = \tau_{a,b}(\vartheta) = \inf\{t \ge \vartheta\colon Y_t \notin (a,b)\}.$$

Then $Y_\rho \sim \nu^{(a,b)}$ under $\mathbb{P}^y$.

Proof. Note that ρ is an $(\mathcal{F}_t)$-stopping time and observe that $\rho = \vartheta$ on $\big\{Y_\vartheta \notin (a,b)\big\}$. Moreover, the strong Markov property of Y implies

$$\mathbb{E}^y[\rho] = \mathbb{E}^y\big[\mathbb{E}^y[\rho \,|\, \mathcal{F}_\vartheta]\big] = \mathbb{E}^y[\vartheta] + \mathbb{E}^y\left[\mathbb{1}_{\{Y_\vartheta \in (a,b)\}}\mathbb{E}^{Y_\vartheta}\big[\inf\{t \in [0,\infty) \colon Y_t \notin (a,b)\}\big]\right]. \qquad \text{(A.2.1)}$$

Proposition 3.1, Chapter VII in [56] states that on $\big\{Y_\vartheta \in (a,b)\big\}$ the mapping $(a,b) \ni x \mapsto \mathbb{E}^x\big[\inf\{t \in [0,\infty) \colon Y_t \notin (a,b)\}\big]$ is bounded. Hence, (A.2.1) entails that $\rho < \infty$, $\mathbb{P}^y$-a.s.

The pathwise continuity of Y implies that $Y_\rho \in \{a,b\}$. Therefore, for $A \in \mathcal{B}(\mathbb{R})$ with $A \subseteq (a,b)$ it holds that

$$\mathbb{P}^y[Y_\rho \in A] = \mathbb{P}^y\big[\{Y_\rho \in A\} \cap \{Y_\vartheta \in (a,b)\}\big] + \mathbb{P}^y\big[\{Y_\rho \in A\} \cap \{Y_\vartheta \notin (a,b)\}\big] = 0.$$

On the other hand, we have for $A \in \mathcal{B}(\mathbb{R})$ with $A \subseteq \mathbb{R}\backslash[a,b]$

$$\mathbb{P}^y[Y_\rho \in A] = \mathbb{P}^y\big[\{Y_\rho \in A\} \cap \{Y_\vartheta \notin (a,b)\}\big] = \mathbb{P}^y[Y_\vartheta \in A] = \nu(A).$$

It remains to show that $\mathbb{P}^y[Y_\rho = a] = \nu^{(a,b)}(\{a\})$. Since $(Y_{t\wedge\vartheta})_{t\in[0,\infty)}$ is bounded, the stopped process $(Y_{t\wedge\rho})_{t\in[0,\infty)}$ is also bounded and thus, $(Y_{t\wedge\rho})_{t\in[0,\infty)}$ is a martingale with respect to $\mathbb{P}^y$ and $(\mathcal{F}_t)$. Dominated convergence, the optional stopping theorem and $\rho < \infty$, $\mathbb{P}^y$-a.s, yield

$$\begin{aligned}\int_{\mathbb{R}} x\,\nu(dx) &= \mathbb{E}^y[Y_\vartheta] = \lim_{t\to\infty} \mathbb{E}^y[Y_{t\wedge\rho\wedge\vartheta}] = \lim_{t\to\infty} \mathbb{E}^y[Y_{t\wedge\rho}] = \mathbb{E}^y[Y_\rho] \\ &= \int_{\mathbb{R}\backslash[a,b]} x\,\nu(dx) + a\mathbb{P}^y[Y_\rho = a] + b\mathbb{P}^y[Y_\rho = b].\end{aligned}$$

In addition,

$$\mathbb{P}^y[Y_\rho \in \{a,b\}] = \mathbb{P}^y\big[Y_\vartheta \in [a,b]\big] = \nu\big([a,b]\big).$$

Therefore,

$$\mathbb{P}^y[Y_\rho = a] = \int_{[a,b]} \frac{b-x}{b-a}\nu(dx) = \nu^{(a,b)}(\{a\}).$$

To sum up, we have shown that $Y_\rho \sim \nu^{(a,b)}$ under $\mathbb{P}^y$. □

In the next lemma we state a sufficient condition guaranteeing that the sequence of stopping times $(\tau_n)_{n\geq N}$ constructed in Algorithm III.3.6 is constant on the event $\{Y_{\tau_N} = d\}$, $d \in J$.

Lemma A.2.2. *Let $\mu \in \mathcal{M}^1$ with $\int_{\mathbb{R}} q_{\bar\mu}(x)\mu(dx) < \infty$. Let $(\mu_n)_{n\in\mathbb{N}}$ be the sequence of probability measures and $(\tau_n)_{n\in\mathbb{N}}$ be the stopping times which are constructed in Algorithm III.3.6. Let $d \in J$. If $u_{\mu_N}(d) = u_\mu(d)$, $\partial_- u_{\mu_N}(d) = \partial_- u_\mu(d)$ and $\partial_+ u_{\mu_N}(d) = \partial_+ u_\mu(d)$ for some $N \in \mathbb{N}$, then on the set $\{Y_{\tau_N} = d\}$ it holds that $\tau_n = \tau_N$, $n \geq N$, and hence $Y_{\tau_n} = d$ for all $n \geq N$.*

Proof. Assume that $u_{\mu_N} \neq u_\mu$. Let $n > N$ such that we choose a tangent t_n to u_μ in x_n in Algorithm III.3.6. Denote by s_n the slope of t_n. We show that $d \notin (a_n, b_n)$, where a_n and b_n are the intersection points of t_n and $u_{\mu_{n-1}}$. Premise that $x_n < d$. Then $s_n > \partial_- u_\mu(d)$ by the concavity of u_μ and the construction of $(u_{\mu_k})_{k\in\mathbb{N}}$. Moreover we have $t_n(d) \geq u_\mu(d) = t_-(d)$, where t_- denotes the left-hand side tangent in d. Assume that $t_n(d) = u_\mu(d) = t_-(d)$, then the potential u_μ is linear on $[x_n, d]$ and $\partial_- u_\mu(d) = s_n$. Thus, it follows that $t_n(d) > t_-(d)$. Since $s_n > \partial_- u_\mu(d) \geq \partial_+ u_\mu(d)$, we conclude that $t_n(x) > t_-(x) \geq u_{\mu_N} \geq u_{\mu_{n-1}}(x)$ for all $x \geq d$. Therefore, the intersection point b_n of t_n and $u_{\mu_{n-1}}$ satisfies $b_n < d$. Accordingly, for the case $x_n > d$ we obtain $a_n > d$. Thus, $d \notin (a_n, b_n)$ and on $\{Y_{\tau_N} = d\}$ the first exit time $\rho(a_n, b_n)$ of the interval (a_n, b_n) equals 0. Hence $\tau_n = \tau_N$ and $Y_{\tau_n} = d$ for all $n \geq N$ on $\{Y_{\tau_N} = d\}$. □

Corollary A.2.3. *Let $\mu \in \mathcal{M}^1$ with $\int_{\mathbb{R}} q_{\bar\mu}(x)\mu(dx) < \infty$. Let $(\mu_n)_{n\in\mathbb{N}}$ be the probability measures and $(\tau_n)_{n\in\mathbb{N}}$ be the stopping times from Algorithm III.3.6. Set $\tau = \lim_{n\to\infty} \tau_n$ and let $d \in J$. If $u_{\mu_N}(d) = u_\mu(d)$, $\partial_- u_{\mu_N}(d) = \partial_- u_\mu(d)$ and $\partial_+ u_{\mu_N}(d) = \partial_+ u_\mu(d)$, then the events $\{Y_\tau = d\}$ and $\{Y_{\tau_n} = d$ for all $n \geq N\}$ only differ by a $\mathbb{P}^{\bar\mu}$-null set.*

Proof. First observe that $\{Y_{\tau_N} = d\} = \{Y_{\tau_n} = d \text{ for all } n \geq N\}$ by Lemma A.2.2. Moreover, we have $Y_{\tau_n} \to Y_\tau$, $\mathbb{P}^{\bar{\mu}}$-a.s as $n \to \infty$ and $Y_\tau \sim \mu$ and $Y_{\tau_n} \sim \mu_n$ under $\mathbb{P}^{\bar{\mu}}$ by Proposition III.3.9. Therefore, it holds that

$$\mu(\{d\}) = \mathbb{P}^{\bar{\mu}}[Y_\tau = d] \geq \mathbb{P}^{\bar{\mu}}[Y_{\tau_n} = d \text{ for all } n \geq N] = \mathbb{P}^{\bar{\mu}}[Y_{\tau_N} = d] = \mu_N(\{d\}).$$

Property 4 of Lemma III.3.2 implies that $\mu_N(\{d\}) = \mu(\{d\})$. Hence $\{Y_\tau = d\}$ and $\{Y_{\tau_n} = d \text{ for all } n \geq N\}$ only differ by a $\mathbb{P}^{\bar{\mu}}$-null set. □

A.3 Appendix: Sequential Testing – Optimal Exit Strategies

We prove two auxiliary results for the proof of Lemma IV.3.1 and Theorem IV.3.2. We first characterize all probability measures $\mu \in \mathcal{A}_3(T,y) \backslash \mathcal{A}_2(T,y)$ with mass points $\alpha, 1-\alpha$ and $b \in (\alpha, 1-\alpha)$ such that $\int_{\mathbb{R}} q_y(x)\mu(dx) = T$.

Lemma A.3.1. *Let $y \in (\alpha, 1-\alpha)$, $T \in \big(0, q(\alpha) - q(y)\big)$ and*

$$\mu^b = \frac{(1-\alpha-b)\big(T+q(y)\big) + (b-y)q(\alpha) - (1-\alpha-y)q(b)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)}\delta_\alpha + \frac{q(\alpha)-T-q(y)}{q(\alpha)-q(b)}\delta_b$$
$$+ \frac{(b-\alpha)\big(T+q(y)\big) - (b-y)q(\alpha) - (y-\alpha)q(b)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)}\delta_{1-\alpha},$$

where $b \in \big(k_y^{-1}(T), g_y^{-1}(T)\big)$ and the functions k_y and g_y are given by (IV.4.5) *and* (IV.4.6), *respectively. Then the measures μ^b, $b \in \big(k_y^{-1}(T), g_y^{-1}(T)\big)$, are exactly all measures in $\mathcal{A}_3(T,y) \backslash \mathcal{A}_2(T,y)$ such that the mass is concentrated in α, b and $1-\alpha$ and $\int_{\mathbb{R}} q_y(x)\mu^b(dx) = T$.*

Proof. Recall that if we fix the three mass points $\alpha, b, 1-\alpha$ for a measure $\mu \in \mathcal{A}_3(T,y) \backslash \mathcal{A}_2(T,y)$, then the constraints $\int_{\mathbb{R}} 1\,\mu(dx) = 1$, $\int_{\mathbb{R}} x\,\mu(dx) = y$ and $\int_{\mathbb{R}} q_y(x)\mu(dx) = T$ uniquely define the weights of a signed measure. Conditions on b then guarantee that we obtain a probability measure.

Let $\mu^b = p_1\delta_\alpha + p_2\delta_b + (1-p_1-p_2)\delta_{1-\alpha}$ be a probability measure with $p_1, p_2, 1-p_1-p_2 \in (0,1)$ and

$$\int_{\mathbb{R}} q_y(x)\mu^b(dx) = (1-p_2)q(\alpha) + p_2 q(b) - q(y) = T < q(\alpha) - q(y).$$

Hence $q(b) < q(\alpha)$, which implies that $b \in (\alpha, 1-\alpha)$ is necessary. Moreover, if we impose that μ^b satisfies $\int_{\mathbb{R}} 1\,\mu^b(dx) = 1$, $\int_{\mathbb{R}} x\,\mu^b(dx) = y$ and $\int_{\mathbb{R}} q_y(x)\mu^b(dx) = T$, then the weights p_1, p_2 and $p_3 = 1-p_1-p_2$ are given by

$$p_1 = \frac{(1-\alpha-b)\big(T+q(y)\big) + (b-y)q(\alpha) - (1-\alpha-y)q(b)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)},$$

$$p_2 = \frac{q(\alpha) - \big(T+q(y)\big)}{q(\alpha)-q(b)},$$

$$p_3 = \frac{(b-\alpha)\big(T+q(y)\big) - (b-y)q(\alpha) - (y-\alpha)q(b)}{(1-2\alpha)\big(q(\alpha)-q(b)\big)}.$$

Using $T < q(\alpha) - q(y)$ and $q(\alpha) > q(b)$ results in $p_2 > 0$. From

$$(1-\alpha-b)\big(T+q(y)-q(\alpha)\big) + (y-\alpha)\big(q(b)-q(\alpha)\big) < 0$$

it follows that $p_1 < 1$. Note that we have $p_2 < 1$ if and only if $q(b) < T + q(y)$. Since $b \mapsto q(b)$ is continuous, symmetric around $\frac{1}{2}$ and strictly decreasing on $\left(\alpha, \frac{1}{2}\right]$ with $q(\alpha) > T + q(y)$ and $q\left(\frac{1}{2}\right) = 0$, there exists a unique $\bar{b} = \bar{b}(T) \in \left(\alpha, \frac{1}{2}\right)$ with $q(\bar{b}) = T + q(y)$ and $q(b) < T + q(y)$ for all

$b \in \left(\bar{b}, 1-\bar{b}\right)$. Thus, we restrict b to the interval $\left(\bar{b}, 1-\bar{b}\right)$. Since $q(y) < T + q(y)$ it follows that $y \in \left(\bar{b}, 1-\bar{b}\right)$. Moreover, it holds that $p_3 > 0$ if and only if

$$\frac{b-y}{b-\alpha} q(\alpha) + \frac{y-\alpha}{b-\alpha} q(b) < T + q(y). \tag{A.3.1}$$

For $b > y$ the Inequality (A.3.1) can be rewritten as $g_y(b) < T$. From the proof of Lemma IV.3.1 we already know that $g_y(b)$ is strictly increasing in b. Hence, for $b > y$ Inequality (A.3.1) holds if and only if $b < g_y^{-1}(T)$. In particular, since for all $b \in \left(y, g_y^{-1}(T)\right)$ we have

$$T > g_y(b) = \frac{b-y}{b-\alpha} q(\alpha) + \frac{y-\alpha}{b-\alpha} q(b) - q(y) > q(b) - q(y),$$

it follows that $g_y^{-1}(T) < 1 - \bar{b}$. For $b \in \left(\bar{b}, y\right]$ we use that

$$T + q(y) > q(b) = \frac{b-y}{b-\alpha} q(\alpha) + \frac{y-\alpha}{b-\alpha} q(b) + \frac{y-b}{b-\alpha}\big(q(\alpha) - q(b)\big) > \frac{b-y}{b-\alpha} q(\alpha) + \frac{y-\alpha}{b-\alpha} q(b).$$

So far we have restricted b to the interval $\left(\bar{b}, g_y^{-1}(T)\right)$. In order to obtain conditions on b which guarantee that $p_1 > 0$, similar arguments as for $p_3 > 0$ apply. We have $p_1 > 0$ if and only if

$$\frac{1-\alpha-y}{1-\alpha-b} q(b) + \frac{y-b}{1-\alpha-b} q(\alpha) < T + q(y). \tag{A.3.2}$$

Using $q(b) < T + q(y)$ yields that (A.3.2) holds for all $b \in \left[y, g_y^{-1}(T)\right)$. For $b < y$ we conclude that (A.3.2) is satisfied if and only if $b > k_y^{-1}(T)$. In addition, (A.3.2) implies that $q(b) < T + q(y)$ for $b \in \left(k_y^{-1}(T), y\right)$ and hence $k_y^{-1}(T) > \bar{b}$. Finally, $p_3 < 1$ if and only if

$$j(b) := (b-\alpha)\big(T + q(y)\big) - (1 - 2\alpha + b - y) q(\alpha) + (1-\alpha-y) q(b) < 0.$$

The function j is convex on $(0,1)$, because $j''(b) = (1-\alpha-y) q''(b) > 0$ with

$$\begin{aligned} j\left(k_y^{-1}(T)\right) &= \frac{1-\alpha-y}{1-\alpha-k_y^{-1}(T)} (1-2\alpha)\big(q(k_y^{-1}(T)) - q(\alpha)\big) < 0, \\ j\left(g_y^{-1}(T)\right) &= (1-2\alpha)\big(q(g_y^{-1}(T)) - q(\alpha)\big) < 0. \end{aligned}$$

Thus, for all $b \in \left(k_y^{-1}(T), g_y^{-1}(T)\right)$ the convexity of j implies that $j(b) < 0$. To summarize, μ^b is a probability measure if and only if $b \in \left(k_y^{-1}(T), g_y^{-1}(T)\right)$. □

We now show that Equation (IV.3.1) posses a unique zero.

Lemma A.3.2. *The equation*

$$(\beta - 1)\big(q(\alpha) - q(b)\big) + (1 - \alpha - b + b\beta - \alpha\beta) q'(b) = 0$$

has a unique solution $b^* \in \left(\alpha, \frac{1}{2}\right]$ *on* $[\alpha, 1-\alpha]$.

Proof. For $b \in (0,1)$ let

$$\ell(b) = (\beta - 1)\big(q(\alpha) - q(b)\big) + (1 - \alpha - b + b\beta - \alpha\beta) q'(b).$$

Observe that ℓ is continuous with $\ell(\alpha) = (1-2\alpha) q'(\alpha) < 0$, $\ell\left(\frac{1}{2}\right) = (\beta - 1) q(\alpha) > 0$ if $\beta > 1$ and $\ell\left(\frac{1}{2}\right) = 0$ if $\beta = 1$. Moreover,

$$\ell'(b) = \big(1 - \alpha - b + \beta(b - \alpha)\big) q''(b) > 0 \tag{A.3.3}$$

for all $b \in [\alpha, 1-\alpha]$. Hence, ℓ is strictly increasing on $[\alpha, 1-\alpha]$. Furthermore, there exists a unique $b^* \in \left(\alpha, \frac{1}{2}\right]$ such that $\ell(b^*) = 0$. □

A.4 Appendix: 2 Points Suffice

Lemma A.4.1. *Let $K \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$, be a set with non-empty interior and $x \in K$ such that there exists a supporting hyperplane at x. Then x is a boundary point.*

Proof. Let $a \in \mathbb{R}^d \setminus \{0\}$ and $b \in \mathbb{R}$ with $a^\top y \geq b$ for all $y \in K$ and $a^\top x = b$. Now consider the set $C := K - x = \{y - x \colon y \in K\}$. Then C has non-empty interior, $0 \in C$, every boundary point of C is uniquely obtained from a boundary point of K and vice versa and there exists a supporting hyperplane to C at 0. More precisely, it holds that $a^\top y \geq 0$ for all $y \in C$. Assume that $0 \notin \partial C$. Then there exists $\varepsilon > 0$ and an ε-neighborhood $B_\varepsilon := \{y \in \mathbb{R}^d \colon |y| < \varepsilon\}$ of 0, which is contained in the interior of C. Let $z \in B_\varepsilon \subseteq C$. Then it also holds true that $-z \in B_\varepsilon$. Hence, we conclude that $a^\top z = 0$ for all $z \in B_\varepsilon$. Now choose the orthonormal basis $e_1 = (1, 0, 0, ..., 0)$, $e_2 = (0, 1, 0, 0, ..., 0), ..., e_d = (0, ..., 0, 1)$ of $\mathbb{R}^d$. Then a can be written as

$$a = \sum_{j=1}^{d} \left(a^\top e_j\right) e_j = \sum_{i=1}^{d} \frac{2}{\varepsilon} \left(a^\top \frac{\varepsilon}{2} e_j\right) e_j = 0,$$

because $\frac{\varepsilon}{2} e_j \in B_\varepsilon$ for all $1 \leq j \leq d$. But $a = 0 \in \mathbb{R}^d$ is not possible. Therefore, 0 is a boundary point of C and, hence, $x \in \partial K$. $\square$

Bibliography

[1] N. I. Achieser. *Theory of approximation.* Dover Publications, Inc., New York, 1992. Translated from the Russian and with a preface by Charles J. Hyman, reprint of the 1956 English translation.

[2] S. Ankirchner, D. Hobson, and P. Strack. Finite, integrable and bounded time embeddings for diffusions. *Bernoulli*, 21(2):1067–1088, 2015.

[3] S. Ankirchner, N. Kazi-Tani, M. Klein, and T. Kruse. Stopping with expectation constraints: 3 points suffice. `http://hal.archives-ouvertes.fr/hal-01525439`, preprint, 2017.

[4] S. Ankirchner, M. Klein, and T. Kruse. A verification theorem for optimal stopping problems with expectation constraints. *Applied Mathematics & Optimization*, pages 1–33, 2017.

[5] S. Ankirchner, M. Klein, T. Kruse, and M. Urusov. On a certain local martingale in a general diffusion setting. `https://hal.archives-ouvertes.fr/hal-01700656`, preprint, 2018.

[6] E. Bayraktar and C. W. Miller. Distribution-constrained optimal stopping. *Mathematical Finance*, pages 1–39, 2018.

[7] E. Bayraktar and S. Yao. Dynamic programming principles for optimal stopping with expectation constraint. *arXiv preprint arXiv:1708.02192*, 2017.

[8] M. Beiglböck, M. Eder, C. Elgert, and U. Schmock. Geometry of distribution-constrained optimal stopping problems. *Probability Theory and Related Fields*, 2018.

[9] R. Bellman. *Dynamic programming.* Dover Publications, Inc., Mineola, NY, 2003. Reprint of the sixth (1972) edition, with an introduction by Eric V. Denardo.

[10] D. Bertsimas and J. N. Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

[11] O. Bokanowski, A. Picarelli, and H. Zidani. State-constrained stochastic optimal control problems via reachability approach. *SIAM J. Control Optim.*, 54(5):2568–2593, 2016.

[12] B. Bouchard, R. Elie, and C. Imbert. Optimal control under stochastic target constraints. *SIAM J. Control Optim.*, 48(5):3501–3531, 2009/10.

[13] B. Bouchard, R. Elie, and N. Touzi. Stochastic target problems with controlled loss. *SIAM J. Control Optim.*, 48(5):3123–3150, 2009/10.

[14] B. Bouchard and M. Nutz. Weak dynamic programming for generalized state constraints. *SIAM J. Control Optim.*, 50(6):3344–3373, 2012.

[15] J. K. Brooks. One-dimensional diffusions and their convergence in distribution. In *Handbook of measure theory, Vol. I, II*, pages 295–343. North-Holland, Amsterdam, 2002.

[16] R. V. Chacon. Potential processes. *Trans. Amer. Math. Soc.*, 226:39–58, 1977.

[17] R. V. Chacon and J. B. Walsh. One-dimensional potential embedding. In *Séminaire de Probabilités, X (Prèmiere partie, Univ. Strasbourg, Strasbourg, année universitaire 1974/1975)*, pages 19–23. Lecture Notes in Math., Vol. 511. Springer, Berlin, 1976.

[18] A. S. Cherny and H.-J. Engelbert. *Singular stochastic differential equations*, volume 1858 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2005.

[19] Y. S. Chow, H. Robbins, and D. Siegmund. *Great expectations: The theory of optimal stopping.* Houghton Mifflin Co., Boston, Mass., 1971.

[20] A. Cox and D. Hobson. An optimal embedding for diffusions. *Stochastic Processes and Applications*, 111:17–39, 2004.

[21] A. M. G. Cox and S. Källblad. Model-independent bounds for Asian options: A dynamic programming approach. *SIAM J. Control Optim.*, 55(6):3409–3436, 2017.

[22] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992.

[23] J. K. Duckworth and M. Zervos. An investment model with entry and exit decisions. *J. Appl. Probab.*, 37(2):547–559, 2000.

[24] R. Durrett. *Probability: Theory and Examples.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[25] J. Elstrodt. *Maß- und Integrationstheorie.* Springer-Lehrbuch. Springer-Verlag, Berlin, 7., korrigierte und aktualisierte Auflage, 2011.

[26] H. J. Engelbert and W. Schmidt. On the behaviour of certain functionals of the Wiener process and applications to stochastic differential equations. In *Stochastic differential systems (Visegrád, 1980)*, volume 36 of *Lecture Notes in Control and Information Sci.*, pages 47–55. Springer, Berlin-New York, 1981.

[27] P. V. Gapeev and G. Peskir. The Wiener sequential testing problem with finite horizon. *Stoch. Stoch. Rep.*, 76(1):59–75, 2004.

[28] X. Guo and L. Shepp. Some optimal stopping problems with nontrivial boundaries for pricing exotic options. *J. Appl. Probab.*, 38(3):647–658, 2001.

[29] V. Henderson, D. Hobson, and A. S. L. Tse. Randomized strategies and prospect theory in a dynamic context. *J. Econom. Theory*, 168:287–300, 2017.

[30] V. Henderson, D. Hobson, and M. Zeng. Optimal stopping and the sufficiency of randomized threshold strategies. *Electron. Commun. Probab.*, 23:Paper No. 30, 11, 2018.

[31] D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. In *Paris-Princeton Lectures on Mathematical Finance 2010*, volume 2003 of *Lecture Notes in Math.*, pages 267–318. Springer, Berlin, 2011.

[32] D. Hobson. Integrability of solutions of the Skorokhod embedding problem for diffusions. *Electron. J. Probab.*, 20:1–26, 2015.

[33] W. Hoeffding. The extrema of the expected value of a function of independent random variables. *Ann. Math. Statist.*, 26:268–275, 1955.

[34] M. Horiguchi. Markov decision processes with a stopping time constraint. *Math. Methods Oper. Res.*, 53(2):279–295, 2001.

[35] P. Jaillet, D. Lamberton, and B. Lapeyre. Variational inequalities and the pricing of American options. *Acta Applicandae Mathematica*, 21(3):263–289, 1990.

[36] S. Källblad. A dynamic programming principle for distribution-constrained optimal stopping. *arXiv preprint arXiv:1703.08534*, 2017.

[37] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[38] S. Karlin and W. J. Studden. *Tchebycheff systems: With applications in analysis and statistics*. Pure and Applied Mathematics, Vol. XV. Interscience Publishers John Wiley & Sons, New York-London-Sydney, 1966.

[39] D. P. Kennedy. On a constrained optimal stopping problem. *J. Appl. Probab.*, 19(3):631–641, 1982.

[40] M. G. Kreĭn. The ideas of P. L. Čebyšev and A. A. Markov in the theory of limiting values of integrals and their further development. *Amer. Math. Soc. Transl. (2)*, 12:1–121, 1959.

[41] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes. I*, volume 5 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, expanded edition, 2001.

[42] C. Makasu. Bounds for a constrained optimal stopping problem. *Optim. Lett.*, 3(4):499–505, 2009.

[43] C. W. Miller. Non-linear PDE Approach to Time-Inconsistent Optimal Stopping. *SIAM Journal on Control and Optimization*, 55(1):557–573, 2017.

[44] P. v. Moerbeke. Optimal stopping and free boundary problems. *Rocky Mountain J. Math.*, 4(3):539–578, 09 1974.

[45] I. Monroe. Using additive functionals to embed preassigned distributions in symmetric stable processes. *Trans. Amer. Math. Soc.*, 163:131–146, 1972.

[46] A. M. Oberman. Wide stencil finite difference schemes for the elliptic Monge-Ampère equation and functions of the eigenvalues of the Hessian. *Discrete Contin. Dyn. Syst. Ser. B*, 10(1):221–238, 2008.

[47] J. Obłój. The Skorokhod embedding problem and its offspring. *Probab. Surveys*, 1:321–392, 2004.

[48] Y. Ogawa. *The Housekeeper and the Professor: A Novel,* translated by Stephen Snyder. Picador, 2009. p. 34.

[49] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Hochschultext / Universitext. Springer, 2003.

[50] J. Palczewski and Ł. Stettner. Infinite horizon stopping problems with (nearly) total reward criteria. *Stochastic Process. Appl.*, 124(12):3887–3920, 2014.

[51] J. Pedersen and G. Peskir. The Azéma-Yor embedding in non-singular diffusions. *Stochastic Processes and Applications*, 96:305–312, 2001.

[52] J. L. Pedersen and G. Peskir. Optimal mean–variance selling strategies. *Math. Financ. Econ.*, 10(2):203–220, 2016.

[53] G. Peskir. A duality principle for the Legendre transform. *J. Convex Anal.*, 19(3):609–630, 2012.

[54] G. Peskir. Optimal detection of a hidden target: The median rule. *Stochastic Process. Appl.*, 122(5):2249–2263, 2012.

[55] G. Peskir and A. Shiryaev. *Optimal stopping and free-boundary problems.* Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2006.

[56] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, third edition, 1999.

[57] R. T. Rockafellar. *Convex Analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[58] L. Rogers and D. Williams. *Diffusions, Markov processes, and Martingales. Vol. 1: Foundations.* Cambridge Mathematical Library. Cambridge University Press, second edition, 2000.

[59] L. Rogers and D. Williams. *Diffusions, Markov processes, and Martingales. Vol. 2: Itô Calculus.* Cambridge Mathematical Library. Cambridge University Press, second edition, 2000.

[60] H. Rost. The stopping distributions of a Markov Process. *Invent. Math*, 14:1–16, 1971.

[61] L. L. Schumaker. *Spline functions: basic theory.* John Wiley & Sons, Inc., New York, 1981. Pure and Applied Mathematics, A Wiley-Interscience Publication.

[62] L. A. Shepp. A first passage problem for the Wiener process. *Ann. Math. Statist.*, 38:1912–1914, 1967.

[63] A. N. Shiryaev. *Optimal Stopping Rules*, volume 8 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin, 2008. Translated from the 1976 Russian second edition by A. B. Aries, reprint of the 1978 translation.

[64] A. V. Skorohod. Исследования по теории случай ных процессов *(*Стохастические дифференциальные уравнения и предельные теоремы для процессов Маркова*).* Izdat. Kiev. Univ., Kiev, 1961.

[65] A. V. Skorokhod. *Studies in the theory of random processes.* Translated from the Russian by Scripta Technica, Inc. Addison-Wesley Publishing Co., Inc., Reading, Mass., 1965.

[66] H. M. Soner and N. Touzi. Stochastic target problems, dynamic programming, and viscosity solutions. *SIAM Journal on Control and Optimization*, 41(2):404–424, 2002.

[67] N. Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29 of *Fields Institute Monographs.* Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2013. With Chapter 13 by Angès Tourin.

[68] M. A. Urusov. On a property of the moment at which Brownian motion attains its maximum and some optimal stopping problems. *Theory of Probability & Its Applications*, 49(1):169–176, 2005.

[69] A. Wald. *Sequential Analysis.* John Wiley & Sons, Inc., New York; Chapman & Hall, Ltd., London, 1947.

[70] D. Williams. *Probability with Martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

[71] G. Winkler. Extreme points of moment sets. *Math. Oper. Res.*, 13(4):581–587, 1988.

[72] Z. Q. Xu and X. Y. Zhou. Optimal stopping under probability distortion. *Ann. Appl. Probab.*, 23(1):251–282, 2013.

[73] X. Yu, C. Zhou, and Y. Zhou. On dynamic programming principle for stochastic control under expectation constraints. *arXiv preprint arXiv:1802.03954*, 2018.

Ehrenwörtliche Erklärung

Hiermit erkläre ich,

- dass mir die Promotionsordnung der Fakultät bekannt ist,
- dass ich die Dissertation selbst angefertigt habe, keine Textabschnitte oder Ergebnisse eines Dritten oder eigener Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel, persönliche Mitteilungen und Quellen in meiner Arbeit angegeben habe,
- dass ich die Hilfe eines Promotionsberaters nicht in Anspruch genommen habe und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten erhalten haben, die in Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen,
- dass ich die Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht habe,
- dass ich bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts nicht von Dritten unterstützt wurde,
- dass ich weder die gleiche noch eine in wesentlichen Teilen ähnliche bzw. eine andere Abhandlung bereits bei einer anderen Hochschule als Dissertation eingereicht habe.

Jena, den 04. April 2018 Maike Klein